JMIR Human Factors

Impact Factor (2022): 2.7
Volume 10 (2023) ISSN: 2292-9495 Editor in Chief: Andre Kushniruk (BA, MSc, PhD, FACMI)

Contents

Original Papers

A Digital Lifestyle Coach (E-Supporter 1.0) to Support People With Type 2 Diabetes: Participatory Development Study (e40017)
Eclaire Hietbrink, Anouk Middelweerd, Pepijn van Empelen, Katharina Preuhs, Annemieke Konijnendijk, Wendy Oude Nijeweme-d’Hollosy, Laura Schrijver, Gozewijn Laverman, Miriam Vollenbroek-Hutten ........................................... 10

Exploring the Use of Pictograms in Privacy Agreements to Facilitate Communication Between Users and Data Collecting Entities: Randomized Controlled Trial (e34855)
Larissa Ugaya Mazza, Laura Fadrique, Amethyst Kuang, Tania Donovska, Hélène Vaillancourt, Jennifer Teague, Victoria Hailey, Stephen Michell, Plinio Morita .......................................................... 36

Development of a Digital Behavioral Intervention to Reduce the Consumption of Sugar-Sweetened Beverages Among Rural Appalachian Adults: Multiphased, Human-Centered Design Approach (e41262)
Jamie Zoellner, Annie Reid, Kathleen Porter, Christina Frederick, Michelle Hilgart, Lee Ritterband .......................................................... 48

Development of a Digital Assistant to Support Teleconsultations Between Remote Physicians and Frontline Health Workers in India: User-Centered Design Approach (e25361)
Neha Verma, Harold Lehmann, Amal Alam, Youseph Yazdi, Soumyadipta Acharya .......................................................... 66

Evaluating Clinician Expectations of mHealth Solutions to Increase Rapid-Screening for HIV and Hepatitis in Migrant Populations in France: Qualitative Study (e41861)
Carter Brown, Guillaume Roucoux, Svetlane Dimi, Saleh Fahmi, Raj-Banou Jeevan, Olivier Chassany, John Chaplin, Martin Duracinsky .... 7

Acceptability of a Health Care App With 3 User Interfaces for Older Adults and Their Caregivers: Design and Evaluation Study (e42145)
Joo Kim, Saguna Saguna, Christer Åhlund .......................................................... 89

Exploring User Visions for Modeling mHealth Apps Toward Supporting Patient-Parent-Clinician Collaboration and Shared Decision-making When Treating Adolescent Knee Pain in General Practice: Workshop Study (e44462)
Simon Johansen, Anne Kanstrup, Kian Haseli, Visti Stenmo, Janus Thomsen, Michael Ratliff .......................................................... 111

Using Intervention Mapping and Behavior Change Techniques to Develop a Digital Intervention for Self-Management in Stroke: Development Study (e45099)
Alex Wong, Mandy Fong, Elizabeth Munsell, Christopher Metts, Sunghoon Lee, Ginger Nicol, Olivia DePaul, Stephanie Tomazin, Katherine Kaufman, David Mohr .......................................................... 132
A Tailored mHealth App for Improving Health and Well-Being Behavioral Transformation in UK Police Workers: Usability Testing via a Mixed Methods Study (e42912)
Richa Mehra, Andy Pulman, Huseyin Dogan, Jane Murphy, Fiona Bitters. .......................................................... 147

Bridging the Communication Gap Between People With Cognitive Impairments and Their Caregivers Using mHealth Apps: User-Centered Design and Evaluation Study With People With 22q11 Deletion Syndrome (e44290)
Martijn Van Dooren, Robin De Croon, Ann Swillen, Katrien Verbert. .......................................................... 164

Acceptance and Usability of an Innovative mDentistry eHygiene Model Amid the COVID-19 Pandemic Within the US National Dental Practice-Based Research Network: Mixed Methods Study (e45418)
Jin Xiao, Dorota Kopycka-Kedziewarski, Patricia Raguza, Luis Mendez Chagoya, Kimberly Funkhouser, Tamara Lisichka, Tong Wu, Kevin Fiscella, Kumari Kar, Nisreen Al Jallad, Noha Rashwan, Johana Ren, Cyril Meyerowitz, National Dental Practice-Based Research Network (PBRN) Collaborative Group. .......................................................... 177

Primary Perspectives in Meme Utilization as a Digital Driver for Medical Community Engagement and Education Mobilization: Pre-Post Study (e40244)
Darrel Wang, Neha Balapal, Amala Ankem, Saishravan Shyamsundar, Adarsh Balaji, Jasmine Kannikal, Marlie Bruno, Shuhan He, Paul Chong. .......................................................... 191

Persuasive Messages for Improving Adherence to COVID-19 Prevention Behaviors: Randomized Online Experiment (e41328)
Mehdi Mourali, Jamie Benham, Raynell Lang, Madison Fullerton, Jean-Christophe Boucher, Kirsten Cornelson, Robert Oxoby, Cora Constantinescu, Theresa Tang, Deborah Marshall, Jia Hu. .......................................................... 201

Health Care Professionals’ Experiences in Telerehabilitation: Qualitative Content Analysis (e40690)
Maria Qvarfordt, Evalill Nilsson, Lina Nilsson. .......................................................... 216

Evaluating the Usability of an Emergency Department After Visit Summary: Staged Heuristic Evaluation (e43729)
Hanna Barton, Megan Salwei, Rachel Rutkowski, Kathryn Wust, Sheryl Krause, Peter Hoonakker, Paula Dail, Denise Buckley, Alexis Eastman, Brad Ehlenfeldt, Brian Patterson, Manish Shah, Barbara King, Nicole Werner, Pascale Carayon. .......................................................... 303

Evaluating the Usefulness and Ease of Use of a Next-Generation–Connected Drug Delivery Device for Growth Hormone Therapy: Qualitative Study of Health Care Professionals’ Perceptions (e46893)
José Labarta, Paul Dimitri, Matthew Keiser, Ekaterina Koledova, Octavio Rivera-Romero. .......................................................... 314

Introduction to the Coproduction of Supervision Standards for Digital Peer Support: Qualitative Study (e40607)
Caroline Collins-Pisano, Michael Johnson, George Mois, Jessica Brooks, Amanda Myers, Deanna Mazina, Marianne Storm, Maggie Wright, Nancy Berger, Ann Kasper, Anthony Fox, Sandi MacDonald, Sarah Schultze, Andrew Bohm, Julia Hill, Karen Fortuna. .......................................................... 329

Perceived Patient Workload and Its Impact on Outcomes During New Cancer Patient Visits: Analysis of a Convenience Sample (e49490)
Safa Elkefi, Onur Asan. .......................................................... 338

Mental Health Professionals’ Attitudes Toward Digital Mental Health Apps and Implications for Adoption in Portugal: Mixed Methods Study (e45949)
Diogo Nogueira-Leite, José Diniz, Ricardo Cruz-Correia. .......................................................... 353

Consumers’ Needs for Laboratory Results Portals: Questionnaire Study (e42843)
Helen Monkman, Janessa Griffith, Leah MacDonald, Blake Lesselroth. .......................................................... 372
Designing and Developing an eHealth Program for Patients With Persistent Physical Symptoms: Usability Study (e42572)
Oliver Christensen, Leonora Hedegaard, Mette Rask, Jane Clemensen, Lisbeth Frosthoml, Marianne Rosendal. ......................................................... 383

Evaluation of an Electronic Care and Rehabilitation Planning Tool With Stroke Survivors With Aphasia: Usability Study (e43861)
Nadia Davood, Aboozar Eghdam, Sabine Koch, Maria Häglund. .......................................................... 395

Trust and Uncertainty in the Implementation of a Pilot Remote Blood Pressure Monitoring Program in Primary Care: Qualitative Study of Patient and Health Care Professional Views (e36072)
Evelyn Chew, Sok Teo, Wern Tang, David Ng, Gerald Koh, Valerie Teo. ......................................................... 414

Characterizing the Gaps Between Best-Practice Implementation Strategies and Real-world Implementation: Qualitative Study Among Family Physicians Who Engaged With Audit and Feedback Reports (e38736)
Geneviève Rouleau, Catherine Reis, Noah Ivers, Laura Desveaux. ......................................................... 429

Patients’ Information Needs Related to a Monitoring Implant for Heart Failure: Co-designed Study Based on Affect Stories (e38096)
Ambre Davat, Fabienne Martin-Juchat. .......................................................... 441

The Acceptability of Technology-Based Physical Activity Interventions in Postbariatric Surgery Women: Insights From Qualitative Analysis Using the Unified Theory of Acceptance and Use of Technology 2 Model (e42178)
Pierre Thériouanne, Meggy Hayotte, Florent Halgand, Fabienne d’Arripe-Longueville. ......................................................... 453

Understanding the Subjective Experience of Long-term Remote Measurement Technology Use for Symptom Tracking in People With Depression: Multisite Longitudinal Qualitative Analysis (e39479)
Katie White, Erin Dawe-Lane, Sara Siddi, Femke Lamers, Sara Simblett, Gemma Riquelme Alacid, Alina Ivan, Inez Myin-Germey, Josep Haro, Carolin Oetzmann, Priya Popat, Aki Rintala, Elena Rubio-Abadal, Til Wykes, Claire Henderson, Matthew Hotopf, Faith Matcham. ......................................................... 465

Pediatric Primary Care Providers’ Perspectives on Telehealth Platforms to Support Care for Transgender and Gender-Diverse Youths: Exploratory Qualitative Study (e39118)
Gina Sequeira, Nicole Kahn, Kevin Bocek, Taranesh Shafii, Peter Asante, Dimitri Christakis, Wanda Pratt, Laura Richardson. ......................................................... 480

Designing a Future eHealth Service for Posthospitalization Self-management Support in Long-term Illness: Qualitative Interview Study (e39391)
Hege Wathe, Ingvild Morken, Marianne Storm, Anne Husebo. .......................................................... 495

Exploring the Cross-cultural Acceptability of Digital Tools for Pain Self-reporting: Qualitative Study (e42177)
Syed Ali, Rebecca Lee, John McBeth, Ben James, Sean McAllister, Alessandro Chiarotto, William Dixon, Sabine van der Veer. ......................................................... 513

Text Messages Exchanged Between Individuals With Opioid Use Disorder and Their mHealth e-Coaches: Content Analysis Study (e37351)
Yerina Ranjit, Warren Davis, Andrea Fentem, Raven Riordan, Rikki Roscoe, Patricia Cavazos-Rehg. ......................................................... 526

eHealth Promoting Stoma Self-care for People With an Elimination Ostomy: Focus Group Study (e39826)
Igor Soares-Pinto, Ana Braga, Isabel Santos, Natália Ferreira, Sandra Silva, Paulo Alves. ......................................................... 537

Experiences of Using an Electronic Health Tool Among Health Care Professionals Involved in Chronic Obstructive Pulmonary Disease Management: Qualitative Analysis (e43269)
André Nyberg, Anna Sondell, Sara Lundell, Sarah Marklund, Malin Tistad, Karin Wadell. ......................................................... 548

Removing Dust From the German Health Care System by Introducing Health Apps Into Standard Care: Semistructured Interview Study (e42186)
Alexandra Heidel, Christian Hagist, Stefan Spinler, Michael Schoeneberger. ......................................................... 563
Optometrists' Perspectives Regarding Artificial Intelligence Aids and Contributing Retinal Images to a Repository: Web-Based Interview Study (e40887)
Aurora Constantin, Malcolm Atkinson, Miguel Bernabeu, Fiona Buckmaster, Baljean Dhillon, Alice McTrusty, Niall Strang, Robin Williams.

Patients With Cardiovascular Disease Revisiting Specialist Physicians via Remote Treatment: Interview Study of Experiences (e43125)
Charlott Ek, Per-Daniel Liljegren, Anette Edin-Liljegren.

Decision Aids for Patients With Head and Neck Cancer: Qualitative Elicitation of Design Recommendations From Patient End Users (e43551)
Eleah Stringer, Julian Lum, Jonathan Livergant, Andre Kushniruk.

Developing Implementation Strategies to Support the Uptake of a Risk Tool to Aid Physicians in the Clinical Management of Patients With Syncope: Systematic Theoretical and User-Centered Design Approach (e44089)
Geneviève Rouleau, Venkatesh Thiruganasambandamoorthy, Kelly Wu, Bahareh Ghaedi, Phuong Nguyen, Laura Desveaux.

Factors Reducing the Use of a Persuasive mHealth App and How to Mitigate Them: Thematic Analysis (e40579)
Markku Kekkonen, Eveliina Korkiakangas, Jaana Laitinen, Harri Oinas-Kukkonen.

Barriers to and Facilitators of the Implementation of Digital Mental Health Interventions as Perceived by Primary Care Decision Makers: Content Analysis of Structured Open-Ended Survey Data (e44688)
Anders Brantnell, Serdar Temiz, Enrico Baraldi, Joanne Woodford, Louise von Essen.

Factors Affecting Digital Tool Use in Client Interaction According to Mental Health Professionals: Interview Study (e44681)
Lauri Lukka, Veli-Matti Karhulahti, J Palva.

Adult Patients’ Experiences of Using a Patient Portal With a Focus on Perceived Benefits and Difficulties, and Perceptions on Privacy and Security: Qualitative Descriptive Study (e46044)
Elisa Son, Eun-Shim Nahm.

Mixed Reality Technology to Deliver Psychological Interventions to Adolescents With Asthma: Qualitative Study Using the Theoretical Framework of Acceptability (e34629)
Kelsey Sharrad, Caitlin Martini, Andrew Tai, Nicola Spurrer, Ross Smith, Adrian Esterman, Ian Gwilt, Debra Sandford, Kristin Carson-Chahhoud.

User-Centered Design of a Digitally Enabled Care Pathway in a Large Health System: Qualitative Interview Study (e42768)
Maggie McCue, Rasha Khatib, Christopher Kabir, Chris Blair, Ben Fehnert, James King, Alexander Spalding, Lara Zaki, Lambros Chrones, Ant Roy, David Kemp.

Designing Virtual Natural Environments for Older Adults: Think-Aloud Study (e40932)
Rikard Lundstedt, Johanna Persson, Carita Häkansson, Susanne Frennert, Mattias Wallergård.

Empowering Researchers to Query Medical Data and Biospecimens by Ensuring Appropriate Usability of a Feasibility Tool: Evaluation Study (e43782)
Christina Schüttler, Maria Zerlik, Julian Gruendner, Thomas Köhler, Lorenz Rosenau, Hans-Ulrich Prokosch, Brita Sedlmayr.

Attitudes Toward the Adoption of 2 Artificial Intelligence–Enabled Mental Health Tools Among Prospective Psychotherapists: Cross-sectional Study (e46859)
Anne-Kathrin Kleine, Eesha Kokje, Eva Lermer, Susanne Gaube.
<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intensive Care Unit Physicians’ Perspectives on Artificial Intelligence–Based Clinical Decision Support Tools: Preimplementation Survey Study (e39114)</td>
<td>Siri van der Meijden, Anne de Hond, Patrick Thoral, EWout Steyerberg, Ilse Kant, Giovanni Cinà, M Arbous</td>
<td>770</td>
</tr>
<tr>
<td>The Perceived Ease of Use and Perceived Usefulness of a Web-Based Interprofessional Communication and Collaboration Platform in the Hospital Setting: Interview Study With Health Care Providers (e39051)</td>
<td>Jason Nie, Christine Heidebrecht, Andrea Zettler, Jacklyn Pearce, Rafael Cunha, Sherman Quan, Elizabeth Mansfield, Terence Tang</td>
<td>782</td>
</tr>
<tr>
<td>The Priorities of End Users of Emergency Department Electronic Health Records: Modified Delphi Study (e43103)</td>
<td>Matthew Yip, Alun Ackery, Trevor Jamieson, Shaun Mehta</td>
<td>796</td>
</tr>
<tr>
<td>Performance of a Web-Based Reference Database With Natural Language Searching Capabilities: Usability Evaluation of DynaMed and Micromedex With Watson (e43960)</td>
<td>Angela Rui, Pamela Garabedian, Marlika Marceau, Ania Syrowatka, Lynn Volk, Heba Edrees, Diane Seger, Mary Amato, Jacob Cambre, Sevan Dulgarian, Lisa Newmark, Karen Nanji, Petra Schultz, Gretchen Jackson, Ronen Rozenblum, David Bates</td>
<td>807</td>
</tr>
<tr>
<td>A Visual Analytic Tool (VIADS) to Assist the Hypothesis Generation Process in Clinical Research: Mixed Methods Usability Study (e44644)</td>
<td>Xia Jing, Vimla Patel, James Cimino, Jay Shubrook, Yuchun Zhou, Brooke Draghi, Mychell Ernst, Chang Liu, Sonsoles De Lacalle</td>
<td>823</td>
</tr>
<tr>
<td>Understanding the Role of Patient Portals in Fostering Interprofessional Collaboration Within Mental Health Care Settings: Mixed Methods Study (e44747)</td>
<td>Keri Durocher, Hwayeon Shin, Brian Lo, Sheng Chen, Clement Ma, Gillian Strudwick</td>
<td>835</td>
</tr>
<tr>
<td>The Polarization of Clinician and Service Staff Perspectives After the Use of Health Information Technology in Youth Mental Health Services: Implementation and Evaluation Study (e42993)</td>
<td>Sarah McKenna, Sarah Piper, William Capon, Alison Crowley, Lucas Lira, Haley LaMonica, Min Chong, Elizabeth Scott, Ian Hickie, Frank Ierlino</td>
<td>847</td>
</tr>
<tr>
<td>Optimizing Patient-Reported Outcome Collection and Documentation in Medical Music Therapy: Process-Improvement Study (e46528)</td>
<td>Samuel Rodgers-Melnick, Seneca Block, Rachael Rivard, Jeffery Dusek</td>
<td>860</td>
</tr>
<tr>
<td>Alerts and Collections for Automating Patients’ Sensemaking and Organizing of Their Electronic Health Record Data for Reflection, Planning, and Clinical Visits: Qualitative Research-Through-Design Study (e41552)</td>
<td>Drashko Nakikj, David Kreda, Nils Gehlenborg</td>
<td>871</td>
</tr>
<tr>
<td>A Web-Based Self-management App for Living Well With Dementia: User-Centered Development Study (e40785)</td>
<td>Abigail Lee, Emese Csipke, Lauren Yates, Esme Moniz-Cook, Orrl McDermott, Steven Taylor, Michael Stephens, Daniel Kelleher, Martin Orrell.</td>
<td>907</td>
</tr>
<tr>
<td>Web-Based COVID-19 Dashboards and Trackers in the United States: Survey Study (e43819)</td>
<td>Melissa Clarkson</td>
<td>926</td>
</tr>
<tr>
<td>The Teach-ABI Professional Development Module for Educators About Pediatric Acquired Brain Injury: Mixed Method Usability Study (e43129)</td>
<td>Lauren Saly, Christine Provvidenza, Hiba Al-Hakeem, Andrea Hickling, Sara Stevens, Lisa Kakonge, Anne Hunt, Sheila Bennett, Rhonda Martinussen, Shannon Scratch</td>
<td>940</td>
</tr>
<tr>
<td>Acceptability and Potential Impact of the #chatsafe Suicide Postvention Response Among Young People Who Have Been Exposed to Suicide: Pilot Study (e44535)</td>
<td>Louise La Sala, Jane Pirkis, Charlie Cooper, Nicole Hill, Michelle Lamblin, Gowri Rajaram, Simon Rice, Zoe Teh, Pinar Thorn, Rifat Zahan, Jo Robinson</td>
<td>940</td>
</tr>
</tbody>
</table>
Usability and Preliminary Efficacy of an Artificial Intelligence–Driven Platform Supporting Dietary Management in Diabetes: Mixed Methods Study (e43959)
Kim Bul, Nikki Holliday, Mohammad Bhuiyan, Cain Clark, John Allen, Petra Wark .......................................................... 956

A Beta-Prototype Chatbot for Increasing Health Literacy of Patients With Decompensated Cirrhosis: Usability Study (e42506)
Jessica Au, Caitlin Falloon, Ayngaran Ravi, Phil Ha, Suong Le .......................................................... 976

eHealth Literacy and Patient Portal Use and Attitudes: Cross-sectional Observational Study (e40105)
Nikita Deshpande, Vineet Arora, Hanna Vollbrecht, David Meltzer, Valerie Press .......................................................... 988

Barriers to Telemedicine Use: Qualitative Analysis of Provider Perspectives During the COVID-19 Pandemic (e39249)
Milan Patel, Hanna Berlin, Abishek Rajkumar, Sarah Krein, Rebecca Miller, Jessie DeVito, Jake Roy, Margaret Punch, Chad Ellimootli, Alex Peahl .......................................................... 994

Investigating the Connections Between Delivery of Care, Reablement, Workload, and Organizational Factors in Home Care Services: Mixed Methods Study (e42283)
Adam Darwich, Anne-Marie Boström, Susanne Guidetti, Jayanth Raghothama, Sebastiaan Meijer .......................................................... 1004

Linking Activity Theory Within User-Centered Design: Novel Framework to Inform Design and Evaluation of Adverse Drug Reaction Reporting Systems in Pharmacy (e43529)
Joel Fossouo Tagne, Reginald Yakob, Rachael Mcdonald, Nilmini Wickramasinghe .......................................................... 1020

Diversity in Stakeholder Groups in Generative Co-design for Digital Health: Assembly Procedure and Preliminary Assessment (e38350)
Pieter Vandekerckhove, Job Timmermans, Antoinette de Bont, Marleen de Mul .......................................................... 1032

An Artificial Therapist (Manage Your Life Online) to Support the Mental Health of Youth: Co-Design and Case Series (e46849)
Aimee-Rose Wrightson-Hester, Georgia Anderson, Joel Dunstan, Peter McEvoy, Christopher Sutton, Bronwyn Myers, Sarah Egan, Sara Tai, Melanie Johnston-Hollitt, Wai Chen, Tom Gedeon, Warren Mansell .......................................................... 1070

Selection of and Response to Physical Activity–Based Social Comparisons in a Digital Environment: Series of Daily Assessment Studies (e41239)
Danielle Arigo, Robert Gray, Diane Dalal, Jennifer Villareale, Jichen Zhu .......................................................... 1093

How Reflective Automated e-Coaching Can Help Employees Improve Their Capacity for Resilience: Mixed Methods Study (e34331)
Aniek Lentferink, Hilbrand Oldenhuis, Hugo Vethuijsen, Lisette van Gemert-Pijnen .......................................................... 1111

Perceptions of a Digital Mental Health Platform Among Participants With Depressive Disorder, Anxiety Disorder, and Other Clinically Diagnosed Mental Disorders in Singapore: Usability and Acceptability Study (e42167)
Ye Phang, Creighton Heaukulani, Wijaya Martanto, Robert Morris, Mian Tong, Roger Ho .......................................................... 1132

Evaluating User Experience With a Chatbot Designed as a Public Health Response to the COVID-19 Pandemic in Brazil: Mixed Methods Study (e43135)

Patient Satisfaction With Speech Recognition in the Exam Room: Exploratory Survey (e42739)
Jeffrey Sippel, Tim Podhajsky, Chen-Tan Lin .......................................................... 1158
Detecting Medication-Taking Gestures Using Machine Learning and Accelerometer Data Collected via Smartwatch Technology: Instrument Validation Study (e42714)
Chrisogonas Odhiambo, Lukacs Ablonczy, Pamela Wright, Cynthia Corbett, Sydney Reichardt, Homayoun Valafar.

Behavioral Predictors of Intention to Use a Text Messaging Reminder System Among People Living With HIV in Rural Uganda: Survey Study (e42952)

Implementation of a Digital Health Tool for Patients Awaiting Input From a Specialist Weight Management Team: Observational Study (e41256)
Petra Hanson, Charlotte Summers, Arjun Panesar, Alexandros Liarakos, Dominic Odouro-Donkor, Danniella Whyte Oshodi, Luke Hailston, Harpal Randeva, Vinod Menon, Michaela de la Fosse, Amit Kaura, Emma Shuttlewood, Mark Lovered, Donna Poole, Thomas Barber.

A Transgender Health Information Resource: Participatory Design Study (e42382)
Brad Morse, Andrey Soares, Bethany Kwan, Marvyn Allen, Rita Lee, Kristen Desanto, Brooke Holliman, Kate Ytell, Lisa Schilling.

Toward the Design of Sensing-Based Medication Adherence Aids That Support Individualized Activities of Daily Living: Survey and Interviews With Patients and Providers (e40173)
Jacob Biehl, Ravi Patel, Adam Lee.

Workload, Usability, and Engagement with a Mobile App Supporting Video Observation of Methadone Take-Home Dosing: Usability Study (e42654)
Bulat Idrisov, Kevin Hallgren, Alyssa Michaels, Sean Soth, James Darnton, Paul Grekin, Steve Woolworth, Andrew Saxon, Judith Tsui.

Digitally Based Blood Pressure Self-Monitoring Program That Promotes Hypertension Self-Management and Health Education Among Patients With Low-Income: Usability Study (e46313)
Jacqueline Poblete, Natalie Vawter, Sydney Lewis, Earl Felisme, Paloma Mohn, Jennifer Shea, Adam Northrup, Jie Liu, Tala Al-Rousan, Job Godino.

Operational Implementation of Remote Patient Monitoring Within a Large Ambulatory Health System: Multimethod Qualitative Case Study (e45166)
Katharine Lawrence, Nina Singh, Zoe Jonassen, Lisa Groom, Veronica Alfaro Arias, Soumik Mandal, Antoinette Schoenthaler, Devin Mann, Oded Nov, Graham Dove.

Accuracy and Reliability of a Suite of Digital Measures of Walking Generated Using a Wrist-Worn Sensor in Healthy Individuals: Performance Characterization Study (e48270)
Nathan Kowahl, Sooyoon Shin, Poulami Barman, Erin Rainaldi, Sara Popham, Ritu Kapur.

Assessing Mood With the Identifying Depression Early in Adolescence Chatbot (IDEABot): Development and Implementation Study (e44388)
Anna Viduani, Victor Cosenza, Helen Fisher, Claudia Buchweitz, Jader Piccin, Rivka Pereira, Brandon Kohrt, Valeria Mondelli, Alastair van Heerden, Ricardo Araújo, Christian Kieling.

Perceived Use Cases, Barriers, and Requirements for a Smart Health-Tracking Toilet Seat: Qualitative Focus Group Study (e44850)
Sander Hermsen, Vera Verbiest, Marije Buiks, Eva Wentink.

Extending the Privacy Calculus to the mHealth Domain: Survey Study on the Intention to Use mHealth Apps in Germany (e45503)
Niklas von Kalkreuth, Markus Feufel.

Usability and Credibility of a COVID-19 Vaccine Chatbot for Young Adults and Health Workers in the United States: Formative Mixed Methods Study (e40533)
Rose Weeks, Pooja Sangha, Lyra Cooper, João Sedoc, Sydney White, Shai Gretch, Assaf Toledo, Dan Lahav, Anna-Maria Hartner, Nina Martin, Jae Lee, Noam Soinim, Naor Bar-Zeev.
A Medical Assistive Robot for Telehealth Care During the COVID-19 Pandemic: Development and Usability Study in an Isolation Ward (e42870)
Ruohan Wang, Honghao Lv, Zhangli Lu, Xiaoyan Huang, Haiteng Wu, Junjie Xiong, Geng Yang. ................................................................. 1350

Eliciting Opinions on Health Messaging During the COVID-19 Pandemic: Qualitative Survey Study (e39697)
Sienna Ruiz, Uzoma Okere, Michelle Eggers, Catina O'Leary, Mary Politi, Fei Wan, Ashley Housten. ............................................................. 1363

The Impact of Individuals' Social Environments on Contact Tracing App Use: Survey Study (e45825)
Atiyeh Sadeghi, Sebastian Pape, David Harborth. ......................................................................................................................... 1379

Design Validation of a Relational Agent by COVID-19 Patients: Mixed Methods Study (e42740)
Ashraful Islam, Beenish Chaudhry. ........................................................................................................................................... 1396

The TeleTriageTeam, Offering Continuity of Personalized Care Through Telemedicine: Development and Evaluation (e46145)
Janneau Claessens, Sigrid Mueller-Schotte, Jeannette van Weerden, Helianthe Kort, Saskia Imhof, Robert Wisse. ......................................................... 1409

The Effects of a Health Care Chatbot's Complexity and Persona on User Trust, Perceived Usability, and Effectiveness: Mixed Methods Study (e41017)
Joshua Biro, Courtney Linder, David Neyens. ........................................................................................................................................... 1421

Development of a Secondary Prevention Smartphone App for Students With Unhealthy Alcohol Use: Results From a Qualitative Assessment (e41088)
Nicolas Bertholet, Elodie Schmutz, John Cunningham, Jennifer McNeely, Gerhard Gmel, Jean-Bernard Daeppen, Véronique Grazioli. ................................................................. 1430

Feasibility and Acceptability of the Aboriginal and Islander Mental Health Initiative for Youth App: Nonrandomized Pilot With First Nations Young People (e40111)
Kylie Dingwall, Josie Povey, Michelle Sweet, Jaylene Friel, Fiona Shand, Nickolai Titov, Julia Wormer, Tamoor Mirza, Tricia Nagel. ................................................................. 1446

A Health App for Evidence-Based Postpartum Information: Development and Validation Study (e38706)
Máyra Silva de Medeiros, Maiara de Oliveira, Leonandro Gurgel, Anna Ribeiro Rodrigues, Maria Micussi, Adriana Magalhães. ........................................................................... 1462

A Technology-Supported Guidance Model to Increase the Flexibility, Quality, and Efficiency of Nursing Education in Clinical Practice in Norway: Development Study of the TOPP-N Application Prototype (e44101)
Andréa Næs, Jaroslav Zlamal, Stijl Linnerud, Simen Steindal, Marianne Solberg. ................................................................................................. 1471

Using Clinical Data Visualizations in Electronic Health Record User Interfaces to Enhance Medical Student Diagnostic Reasoning: Randomized Experiment (e38941)
Lucille Cheng, Yalini Senathirajah. ............................................................................................................................................... 1489

User Intentions to Use ChatGPT for Self-Diagnosis and Health-Related Purposes: Cross-sectional Survey Study (e47564)
Yeganeh Shahsavar, Avishek Choudhury. ................................................................................................................................. 1498

Reviews

Involving Older People With Frailty or Impairment in the Design Process of Digital Health Technologies to Enable Aging in Place: Scoping Review (e37785)
Emilie Wegener, Jenny Bergschöld, Carly Whitmore, Marjolein Winters, Lars Kayser. ................................................................................................. 228
Exploring Patient Journey Mapping and the Learning Health System: Scoping Review (e43966)
Amanda Joseph, Helen Monkman, Andre Kushniruk, Yuri Quintana. ................................................................. 242

Assessing the Quality and Impact of eHealth Tools: Systematic Literature Review and Narrative Synthesis (e45143)
Christine Jacob, Johan Lindeque, Alexander Klein, Chris Ivory, Sabina Heuss, Marc Peter. ........................................ 251

Feasibility and Acceptability of Chatbots for Nutrition and Physical Activity Health Promotion Among Adolescents: Systematic Scoping Review With Adolescent Consultation (e43227)
Rui Han, Allyson Todd, Sara Wardak, Stephanie Partridge, Rebecca Raeside. ....................................................... 272

Factors That Affect Knowledge-Sharing Behaviors in Medical Imaging Departments in Cancer Centers: Systematic Review (e44327)
Maryam Almashmoum, James Cunningham, Ohoud Alkhaldi, John Anisworth. .................................................... 284

Involving Health Care Professionals in the Development of Electronic Health Records: Scoping Review (e45598)
Theresa Busse, Chantal Jux, Johannes Laser, Peter Rasche, Horst Vollmar, Jan Ehlers, Sven Kernebeck. .................. 1049

Corrigenda and Addenda

Correction: Performance of a Web-Based Reference Database With Natural Language Searching Capabilities: Usability Evaluation of DynaMed and Micromedex With Watson (e48468)
Angela Rui, Pamela Garabedian, Marlika Manceau, Ania Syrowatka, Lynn Volk, Heba Edrees, Diane Seger, Mary Amato, Jacob Cambre, Sevan Dulgarian, Lisa Newmark, Karen Nanji, Petra Schulz, Gretchen Jackson, Ronen Rozenblum, David Bates. ......................... 1091
A Digital Lifestyle Coach (E-Supporter 1.0) to Support People With Type 2 Diabetes: Participatory Development Study

Eclaire A G Hietbrink\textsuperscript{1,2}, MSc; Anouk Middelweerd\textsuperscript{1}, PhD; Pepijn van Empelen\textsuperscript{3}, PhD; Katharina Preuhs\textsuperscript{3}, MSc; Annemieke A J Konijnendijk\textsuperscript{1}, PhD; Wendy Oude Nijeweme-d’Hollosy\textsuperscript{1}, PhD; Laura K Schrijver\textsuperscript{3}, MSc; Gozewijn D Laverman\textsuperscript{1,4}, MD, PhD; Miriam M R Vollenbroek-Hutten\textsuperscript{1,5}, PhD

\textsuperscript{1}Department of Biomedical Signals and Systems, University of Twente, Enschede, Netherlands
\textsuperscript{2}ZGT Academy, Ziekenhuisgroep Twente, Almelo, Netherlands
\textsuperscript{3}Department of Child Health, TNO (Netherlands Organization for Applied Scientific Research), Leiden, Netherlands
\textsuperscript{4}Department of Internal Medicine/Nephrology, Ziekenhuisgroep Twente, Almelo, Netherlands
\textsuperscript{5}Board of Directors, Medisch Spectrum Twente, Enschede, Netherlands

Corresponding Author:
Eclaire A G Hietbrink, MSc
Department of Biomedical Signals and Systems
University of Twente
Drienerlolaan 5
Enschede, 7500 AE Enschede
Netherlands
Phone: 31 53 4892760
Email: e.a.g.hietbrink@utwente.nl

Abstract

Background: A healthy lifestyle, including regular physical activity and a healthy diet, is becoming increasingly important in the treatment of chronic diseases. eHealth interventions that incorporate behavior change techniques (BCTs) and dynamic tailoring strategies could effectively support a healthy lifestyle. E-Supporter 1.0 is an eCoach designed to support physical activity and a healthy diet in people with type 2 diabetes (T2D).

Objective: This paper aimed to describe the systematic development of E-Supporter 1.0.

Methods: Our systematic design process consisted of 3 phases. The definition phase included the selection of the target group and formulation of intervention objectives, and the identification of behavioral determinants based on which BCTs were selected to apply in the intervention. In the development phase, intervention content was developed by specifying tailoring variables, intervention options, and decision rules. In the last phase, E-Supporter 1.0 integrated in the Diameter app was evaluated using a usability test in 9 people with T2D to assess intervention usage and acceptability.

Results: The main intervention objectives were to stimulate light to moderate-vigorous physical activities or adherence to the Dutch dietary guidelines in people with T2D. The selection of behavioral determinants was informed by the health action process approach and theories explaining behavior maintenance. BCTs were included to address relevant behavioral determinants (eg, action control, self-efficacy, and coping planning). Development of the intervention resulted in 3 types of intervention options, consisting of motivational messages, behavioral feedback, and tailor-made supportive exercises. On the basis of IF-THEN rules, intervention options could be tailored to, among others, type of behavioral goal and (barriers to) goal achievement. Data on these variables could be collected using app data, activity tracker data, and daily ecological momentary assessments. Usability testing revealed that user experiences were predominantly positive, despite some problems in the fixed delivery of content.

Conclusions: The systematic development approach resulted in a theory-based and dynamically tailored eCoach. Future work should focus on expanding intervention content to other chronic diseases and lifestyle behaviors, enhancing the degree of tailoring and evaluating intervention effects on acceptability, use, and cost-effectiveness.

\textcopyright JMIR Hum Factors 2023;10:e40017 doi:10.2196/40017
KEYWORDS

eHealth; mHealth; diet; nutrition; physical activity; lifestyle change; coaching; dynamic tailoring; behavior change; blended care; type 2 diabetes; design; treatment; chronic disease; behavioral; theory; intervention; acceptability; usability; cost

Introduction

Background

In 2020, nearly 60% of Dutch adults had experienced one or more chronic diseases [1]. Treatment predominantly focused on drug therapies to make the disease manageable. Although prescribing medication is often an important treatment option, it does not address the fact that chronic diseases can be exacerbated by an unhealthy lifestyle [2,3]. It has been increasingly recognized that a healthy lifestyle, such as a healthy diet, adequate physical activity, and enough sleep, can contribute to a reduction in disease burden, improved quality of life (QoL), and reversal of chronic diseases [4-7]. For example, several studies have shown positive effects of a healthy lifestyle on glycemic control, QoL, medication use, and risk of complications in people with type 2 diabetes (T2D) [8-14]. Furthermore, in people with chronic pulmonary diseases, a healthy lifestyle has shown to improve QoL and to reduce hospitalizations and mortality [15].

Despite the positive effects of healthy lifestyle on the course of chronic diseases, people often find it challenging to live healthy. For instance, >50% of Dutch adults adhered to the Dutch Physical Activity Guidelines [16] and only a quarter of Dutch adults met the Dutch dietary guidelines [17]. Moreover, adherence to healthy lifestyle behaviors appears to be even lower among people with chronic diseases [18-21]. A higher adherence to these guidelines is important for the positive effects of a healthy lifestyle to reach its potential. Realizing sustainable lifestyle improvements requires individuals to self-manage their behavior using a range of skills, such as knowledge acquisition, self-monitoring, action, and coping planning [22]. Developing self-management skills has proven successful in allowing individuals to effectively manage their disease and improve health outcomes [23,24]. However, self-management skills are often insufficient in people with chronic diseases, especially regarding lifestyle behaviors [25-27]. Self-management skills vary from person to person and are subject to various factors (eg, health literacy and socioeconomic status) [28], resulting in each individual needing a personal approach. Hence, interventions with extensive guidance and more motivational strategies fitting the individual's characteristics and needs are required to achieve sustainable lifestyle behavior change. However, it is very costly and challenging to provide extensive guidance via face-to-face programs alone with the rising capacity issues and limited financial resources available in health care domain [29,30]. Therefore, eHealth (ie, the use of technology to support health, well-being, and health care [31]) can be used to contribute to continuous and affordable lifestyle self-management support for people with chronic diseases [32,33].

eHealth has the potential to support lifestyle self-management [34,35]. Guided eHealth interventions have been shown to be as effective as face-to-face treatment in the short term but are generally more cost-effective [36-38]. Moreover, eHealth enables continuous support in daily life and tailoring of support toward an individual [33,39]. Several reviews and meta-analyses showed that app-based eHealth interventions have the potential to improve physical activity levels [40,41] and adherence to dietary guidelines [42,43], resulting in improved health outcomes, such as perceived fitness, body weight, blood pressure, or glycemic control [42-46]. However, eHealth effectiveness has been shown to differ between interventions [47,48], and positive intervention effects are often not sustained in the long term [49-53].

One source of variability in eHealth effectiveness is differences in the use of behavior change theory [54-56]. Behavior change theories have been developed to explain health behaviors and guide health behavior change based on a variety of factors that individually influence and affect the performance of health behavior (ie, behavioral determinants) [57]. Using theory helps us to identify behavioral determinants of behavior that are relevant to the target by means of an intervention to effectively change behavior. Besides, it enables us to determine which behavior change techniques (BCTs) are most likely to bring about change [58,59]. Including behavior change theory is key as interventions have shown to be more effective when they are theory based [46,60-63]. Moreover, using theories specifically explaining behavioral maintenance, in addition to theories that primarily focus on initial change, may be useful for the design of interventions to achieve sustainable behavior change [64,65].

Systematic reviews showed that tailored eHealth interventions are more effective in promoting healthy behaviors and user engagement than generic interventions [44,47,63,66-68]. The effect sizes of static tailored interventions (ie, coaching based on a single baseline assessment) remain small, whereas dynamically tailored interventions (ie, coaching based on iterative assessment) show larger effect sizes and have long-term effects [69,70]. Smartphones and activity trackers enable us to dynamically tailor interventions [71] to an individual (eg, personal goals and self-efficacy levels) and their environment (eg, location and weather) [72]. Given the positive attitude toward smartphone and technology use by both the young and older adults [73,74], it is assumed that dynamically tailored eHealth interventions provide a great opportunity to facilitate behavior change.

eHealth interventions are expected to be more beneficial when provided in a blended care setting (ie, combining eHealth and regular health care) [75-78]. A previous study has shown that blended care, including personal guidance of patients, leads to a higher and better use of eHealth interventions [75], which can result in improved intervention effects [31]. Moreover, the first studies in this emerging research field showed positive effects on intervention effectiveness [36,79-81]. Besides, it is foreseen that a good integration of eHealth and regular care can lead to higher quality of care and a decrease in consultations with health care providers [82].

These studies have shown that dynamic tailoring of interventions can improve health outcomes, such as physical activity levels [79] and adherence to dietary guidelines [80], resulting in improved health outcomes, such as perceived fitness, body weight, blood pressure, or glycemic control [81,82]. However, eHealth effectiveness has been shown to differ between interventions [83,84], and positive intervention effects are often not sustained in the long term [85-87].

One source of variability in eHealth effectiveness is differences in the use of behavior change theory [88-90]. Behavior change theories have been developed to explain health behaviors and guide health behavior change based on a variety of factors that individually influence and affect the performance of health behavior (ie, behavioral determinants) [91]. Using theory helps us to identify behavioral determinants of behavior that are relevant to the target by means of an intervention to effectively change behavior. Besides, it enables us to determine which behavior change techniques (BCTs) are most likely to bring about change [92,93]. Including behavior change theory is key as interventions have shown to be more effective when they are theory based [94,95-98]. Moreover, using theories specifically explaining behavioral maintenance, in addition to theories that primarily focus on initial change, may be useful for the design of interventions to achieve sustainable behavior change [99,100].
care professionals [75,82] and can be more cost-effective than regular care [83].

**Objective**

In summary, innovative eHealth interventions that are theory based, which include dynamic tailoring, and are offered as blended care may help to effectively support health behavior change. However, eHealth interventions that integrate these potential success factors are scarce to date. Therefore, the aim of this paper is to successively describe the systematic development and usability testing of the first version of E-Supporter, a theory-based, dynamically tailored, and blended lifestyle coaching intervention for people with chronic diseases.

**Methods**

**Ethics Approval**

Ethics approval was obtained from the Medical Research Ethics Committees United Nieuwegein, the Netherlands (R20.121). Written consent was requested from each patient to participate in the study. All participants gave verbal consent before starting the audio recording of the interviews. Data privacy was protected in accordance with the General Data Protection Regulation standard. Participants were informed in detail about how data were collected, processed, and stored in the subject information sheet. Participants gave explicit consent for the use of their data by signing the informed consent form. Data privacy was protected by offering anonymous preset accounts without personal data to prevent sharing personal data with commercial parties.

**Project Overview**

The development of E-Supporter is part of the eManager project [84], which aims to enhance patient-centered health care to reduce disease burden in the Netherlands. The eManager project focuses on blended coaching, which consists of the Assessment of Burden of Chronic Conditions (ABCC) tool and E-Supporter (Figure 1).

The ABCC tool is already being used during consultations with a health care professional and is developed for several chronic diseases, including chronic obstructive pulmonary disease, asthma, T2D, and chronic heart failure. The ABCC tool maps the patient’s disease burden based on a questionnaire that the patient completes before consultation and based on medical information from the electronic patient file [85]. During consultation, the burden of disease in various generic and disease-specific domains is clearly presented in a balloon chart (Figure 2), which shows domains that can be improved. The ABCC tool also shows treatment recommendations based on existing guidelines for each domain when selecting a balloon from the image. On the basis of a discussion between the patient and health care professional after the treatment advice, personalized treatment goals can be determined. Detailed information on the development of the ABCC tool can be found elsewhere [85].

**Figure 1.** eManager chronic diseases. ABCC: Assessment of Burden of Chronic Conditions.
E-Supporter aims to support the patient in daily life in pursuing the earlier established personalized treatment goals. Diabetes professionals can offer E-Supporter to patients in a shared decision-making process. E-Supporter focuses on obtaining and maintaining a healthy lifestyle to reduce the perceived burden of disease. In this paper, the development of the first version of E-Supporter (ie, the components as presented within the Monitoring and Coaching frame; Figure 1) is described. In future versions, E-Supporter will also be used to guide health care professionals to be better informed about their patients based on the progress that can be monitored by means of E-Supporter.

Intervention Development Approach

Overview

The systematic design of the intervention was guided by program-planning models [39,86], and the design was categorized into 3 phases: (1) a definition phase in which we analyzed the program goals and possible behavior change strategies, (2) a development phase for which we relied on Nahum-Shani’s model [39] for describing the elements of the intervention, and (3) the evaluation phase in which the developed content was evaluated (Textbox 1). Content for specific target groups or lifestyle behaviors can be added step by step by repeating the development process.

We followed a participatory development process by combining the perspectives of 9 health care professionals and 33 people with T2D. The team of health care professionals consisted of 4 internists, 3 diabetes nurses, and 2 physician researchers. The health care professionals contributed during all phases of the development approach at research group meetings and focus groups by formulating intervention objectives, formulating lifestyle advice for people with T2D, and reviewing and revising the motivational message. Multimedia Appendix 1 provides detailed information on the methods for the focus group discussions for assessing and revising the motivational messages.

The requirements for digital lifestyle coaching from the patient’s perspective emerged from interviews with 19 people with T2D. The results of these interviews are described elsewhere [87]. These requirements were translated into content for E-Supporter 1.0. During the development process, subsets of the motivational messages were evaluated in iterations among 14 people with T2D to gain insights into their opinions. The input derived from these iterative evaluations was used to refine and improve the intervention content.
Description of the stepwise process for the development of E-Supporter.

**Phase 1: Definition**

The program-planning model developed by Kreuter et al [86,88] provided a basis for the step-by-step plan to define the theoretical framework of the intervention. First, our choice to select people with T2D as the target group and physical activity and nutrition as target behaviors was substantiated. Second, the intervention objectives were determined based on the existing Dutch standards of care for people with T2D and guidelines regarding physical activity and healthy nutrition. Third, the selection of changeable behavioral determinants that needed to be addressed in the intervention was guided by behavioral theories. Finally, there was the need to identify BCTs to influence these determinants. We searched in literature for BCTs that could be linked to a particular determinant or the selected health behaviors.

**Phase 2: Development**

In accordance with Nuham-Shani’s model [39], the requirements from the definition phase were translated into intervention content by defining (1) tailoring variables that comprise the information that is used to decide when and how to intervene, (2) intervention options that are defined as the type of support offered (eg, information, feedback, and advice), and (3) decision rules, including the operationalization of decision points when a particular intervention option is delivered. The last step of the development phase comprised the integration of E-Supporter 1.0 in two mobile health apps: (1) the Diameter app [89,90] and (2) MiGuide [91]. Both apps provide blended lifestyle support for people with T2D by means of lifestyle monitoring and coaching.

**Phase 3: Evaluation Phase**

After being integrated in the Diameter app, E-Supporter 1.0 was evaluated during a 5-week usability study among 9 people with T2D. The Diameter app was used because its purpose closely aligned with the initial design and application of E-Supporter 1.0 to encourage physical activity and a healthy diet in people with T2D. This study was the first to evaluate the E-Supporter integrated into an app. Therefore, the aim of the study was two-fold: (1) to gain insight into intervention use and acceptability of E-Supporter 1.0 integrated in the Diameter app.
and (2) to identify technical issues in the integration of E-Supporter 1.0 within the Diameter app. In total, 9 patients with T2D visiting the outpatient clinic at the Ziekenhuis Groep Twente Hospital were recruited; they were ≥18 years and were familiar with an Android smartphone (version 5.0 or higher). Patients were not able to participate when they underwent renal replacement therapy, were engaged in drug abuse, or had insufficient proficiency in the Dutch language. Participants used the Diameter app, with E-Supporter content, in combination with the activity tracker Fitbit Inspire 2 [92] and Freestyle Libre 2 sensor (ie, a continuous glucose monitoring sensor in the interstitial fluid of the upper arm) [93,94] for 5 weeks at home. The Freestyle Libre 2 sensor was one of the self-monitoring tools of the Diameter app to provide continuous insights into glucose values for people with T2D. Participants were asked to use the Diameter app as instructed by the researchers. For this, participants were asked to scan the Freestyle Libre 2 sensor at least 3 times a day (to prevent data loss), to wear the Fitbit activity tracker every day, and to fill in the food diary for at least 6 days [95]. As part of the E-Supporter 1.0 components, participants had the opportunity to set a lifestyle goal, read daily motivational messages, and perform weekly psychological exercises.

A mixed methods approach was used to explore intervention use and acceptability of E-Supporter 1.0 and its integration within the Diameter app. Intervention use was exploratively assessed using log data of the E-Supporter 1.0 components (ie, motivational messages and physiological exercises) and Diameter app components (ie, Fitbit, food diary and Freestyle Libre 2). Use of the Diameter app and E-Supporter components was reported by describing frequency and duration of the used functionalities. Log data were also used to identify whether the intervention was delivered as intended and to identify technical integration issues. Open-ended interviews based on the Unified Theory of Acceptance and Use of Technology 2 model [96] were conducted to capture participants’ experiences with E-Supporter integrated into the Diameter app. Interview topics included the general appreciation of the Diameter; ease of use of the Diameter; perceived usefulness of the Diameter; perceived usefulness of the E-Supporter content; appreciation and perceived enjoyment of the E-Supporter content; technical infrastructure to use the Diameter; and the Diameter in the health care process. The transcripts were coded using inductive thematic analysis [97]. Multimedia Appendix 2 provides more detailed information about the methods of the usability study.

Results

Phase 1: Definition Phase

Step 1: Target Population and Behaviors

E-Supporter 1.0 focused on improving physical activity and a healthy diet in people with T2D. Diabetes is one of the four major types of noncommunicable diseases worldwide [98]. It is also expected that until 2040 the prevalence of diabetes will rise relatively sharply compared with other diseases. This prospect calls for initiatives to reduce the burden of T2D on patients and the health care system. There is growing evidence that lifestyle interventions can positively contribute to the management of T2D [6]. Many studies showed that lifestyle interventions targeting physical activity or diet can achieve reversion or remission of T2D [8,9,11-13]. The aforementioned studies provide insight into the importance of sufficient physical activity and a healthy diet for improved glycemic regulation, reduction of medication use, and possible reversal of T2D. However, a substantial proportion of people with T2D do not meet physical activity and eating guidelines [21,99-103]. To illustrate, previous studies found that adherence to Dutch Healthy Diet Guidelines was low among participants with T2D [102]. In another example, more than one-third of the participants with T2D had limited physical activity (ie, <5000 steps per day) [103]. Therefore, it is important to improve these behaviors among people with T2D so that they can benefit from the positive effects on their health.

Step 2: Intervention Objectives

Physical Activity

The Dutch health care standard for T2D states that small improvements in physical activity levels can already lead to positive health effects, although being physically active for longer periods more often and more intensively does have additional health benefits [104]. The Dutch Health Council emphasizes the aim to achieve the following physical activity levels for the general population: 150 minutes (about 2.5 hours) per week of moderate to vigorous physical activity spread over several days, muscle- and bone-strengthening activities at least twice a week, and to minimize sitting hours [105]. Moreover, it is stated that advice should be aligned with the motivation, possibilities, and daily routine of the person with T2D. Current physical activity guidelines make no statements about light-intensity physical activities (ie, activities that are classified as >1.5 to <3 metabolic equivalents [106], such as slow walking, shopping, or household chores). However, recent studies have shown that light physical activities are beneficially associated with health outcomes [107,108]. People may be more inclined to replace physical inactivity with light-intensity physical activities, which are usually easier to incorporate into daily life [109]. Therefore, the primary aim was to facilitate small step-by-step improvements or maintenance of light to moderate-vigorous physical activities in people with T2D, aligning with their motivation, possibilities, and daily routine.

Nutrition

The Dutch Guidelines for a good diet from 2015 were used to develop the nutritional module of E-Supporter [104,110]. The Dutch Health Council states that essential elements for a healthy diet for people with T2D are already part of the national guidelines, such as reducing the consumption of unhealthy carbohydrate-rich foods (eg, refined grain products). In line with the Dutch Health Council’s recommendations, the main goal of the nutritional module was to increase adherence to the Dutch dietary guidelines, again matching the needs of the individuals closely.

Step 3: Selection of Behavioral Determinants

The health action process approach (HAPA) [111] and theories that elucidate behavioral maintenance (eg, Rothman’s theory of maintenance [112,113] or Marlatt’s relapse prevention theory
[114]) were used as a basis for E-Supporter 1.0. The HAPA model distinguishes between a preintentional motivation phase and a postintentional volition phase each with different behavioral determinants [111]. The HAPA model has shown to be able to explain several health behaviors, including physical activity and dietary behaviors. The volitional phase in the HAPA model comprises both action initiation and maintenance [115]; so, it does not include a separate phase to address behavior maintenance. Because health behavior change is often not maintained in the long term [49-53], addressing behavioral maintenance as well after initial change is emphasized [112]. Behavioral maintenance theories explicitly address determinants important for maintenance of behavior in the long term, such as the formation of habits and the perceived value of a new behavior [113]. Because these theories suggest that different behavioral determinants contribute to behavioral initiation and maintenance, it can be argued that separate determinants targeting behavior maintenance could also be included in the intervention approach. As a result, we developed an intervention approach that recognizes three distinct phases of behavior change: (1) an initiation phase to form intentions to adopt a healthy behavior, (2) an action phase to transform intentions into actual behavior change, and (3) a maintenance phase to support persistence of behavior change in the long term (Figure 3). Determinants of behavior were extracted from the HAPA model and behavior maintenance theories. These determinants were addressed in all phases (ie, key determinants) or in one of the three behavioral phases (ie, phase-specific determinants).

Figure 3. Intervention targets of E-Supporter 1.0.

<table>
<thead>
<tr>
<th>All phases</th>
<th>Key determinants</th>
<th>Self-efficacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initiation phase</td>
<td>Knowledge</td>
<td>Risk perception</td>
</tr>
<tr>
<td>Action phase</td>
<td>Action planning</td>
<td>Coping planning</td>
</tr>
<tr>
<td>Maintenance phase</td>
<td>Habits</td>
<td>Satisfaction</td>
</tr>
</tbody>
</table>

In total, 2 key determinants were selected that recur in all phases of the intervention: action control and self-efficacy. Action control comprises three self-regulatory processes: (1) awareness of standards (ie, a self-set goal), (2) self-monitoring that yields information about the attainment of individual’s behavior or goal, and (3) self-regulatory effort to achieve the goal [116,117]. Self-efficacy refers to an individual’s belief in his or her own capability to perform a certain behavior needed to achieve a desired outcome [118]. Self-efficacy is found to be related to the intention to change [119], goal level and goal achievement, and affective reactions, which have an impact on self-regulatory processes that subsequently influence performance of the target behavior [120]. In the initiation phase, intervention options focus primarily on confidence in one’s own capacity to perform the desired behavior (ie, task self-efficacy). Later in the process, this focus shifts to confidence in one’s own capacity to deal with barriers (ie, maintenance self-efficacy) or to recover from setbacks (ie, recovery self-efficacy) [111].

Phase-specific determinants were derived for each of the defined phases of behavior change (Figure 3). To form intentions for behavior change, the determinants risk perception, outcome expectancies, attitude, and social support were selected from the HAPA model and Rothman’s theory. Moreover, knowledge about healthy behavior within the target group is insufficient [121-123] but is required to achieve self-management to realize lifestyle changes [124]. Therefore, the initiation phase was supplemented with the determinant knowledge, for example, by providing general information about guidelines for physical activity and healthy nutrition. Action planning and coping planning were extracted from the HAPA model as phase-specific determinants in the action phase to translate intention into behavior. For the maintenance phase, we focused on the determinants habits, satisfaction, social influences, and mood regulation. The importance of forming habits and satisfaction with behavior were derived from Rothman’s theory of behavior maintenance. In addition, several habit theories emphasize the role of social influences on behavior maintenance [65]. Social influences can increase an individual’s capacity to maintain behavior because social influences can affect an individual’s opinions, emotional states, and behaviors in the long term. Finally, mood regulation was targeted in our intervention. Marlatt and Gordon [114,125] argued that relapse prevention is an important part of coping planning that refers to not only behavioral adaptation but also mood management or repair in case of a behavioral lapse. An overview of the definitions of all determinants can be found in Multimedia Appendix 3 [126-135].

**Step 4: Selection of BCTs**

The selected BCTs per determinant based on the literature search can be found in Table 1.
We covered 3 BCTs identified by Abraham and Michie that address self-regulatory processes from the concept of negative feedback control [59,116], including review of goals, feedback on behavior, and self-monitoring of behavior. Several studies provided evidence for the effectiveness of BCTs that address self-regulatory processes in changing health behavior [136-139]. To illustrate, Michie et al [139] found that interventions that included self-monitoring of behavior in combination with at least one other self-regulatory technique (eg, goal setting, feedback on behavior, or review of behavioral goals) were significantly more effective in promoting physical activity and healthy eating compared with interventions which did not include these techniques. Regarding self-efficacy, we incorporated BCTs recognized as effective to increase self-efficacy in literature. Systematic reviews and meta-analyses [140,141] found that several BCTs were significantly associated with improvements in self-efficacy levels and positive changes in physical activity. Of these BCTs, action planning, social support, and instruction on how to perform the behavior were included in the intervention. Furthermore, BCTs were derived to target self-efficacy from 2 studies that identified effective BCTs for several prominent determinants of behavior. First, Kok et al [142] described BCTs that target determinants of behavior based on literature synthesis. Second, Johnston et al [143] linked BCTs to determinants of behavior through the triangulation of findings from literature synthesis and expert consensus. The following BCTs were selected from the aforementioned studies to target self-efficacy levels: problem solving, verbal persuasion about capabilities, focus on past success, reduce negative emotions, goal setting, and self-monitoring of behavior.

Table 1. Determinants and linked behavior change techniques incorporated in E-Supporter 1.0.

<table>
<thead>
<tr>
<th>Action control</th>
<th>Self-efficacy</th>
<th>Knowledge</th>
<th>Risk perception</th>
<th>Outcome expectations</th>
<th>Attitude</th>
<th>Social support</th>
<th>Action planning</th>
<th>Coping planning</th>
<th>Mood management</th>
<th>Habits</th>
<th>Satisfaction</th>
<th>Social influences</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Goal-setting (behavior)</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.2 Problem-solving</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4 Action planning</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.5 Review behavior goal</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.2 Feedback on behavior</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.3 Self-monitoring of behavior</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.4 Self-monitoring of outcomes of behavior</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.1 Social support, including motivational interviewing</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.2 Social support (practical)</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.1 Instruction on how to perform the behavior</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.1 Information about health consequences</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.6 Information about emotional consequences</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.3 Information about others’ approval</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.1 Prompts and cues</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.3 Habit formation</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.1 Credible source</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.2 Pros and cons</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.3 Comparative imagining of future outcomes</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.2 Reduce negative emotions</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13.2 Framing and reframing</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.1 Verbal persuasion about capability</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.3 Focus on past success</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
BCTs that previously were shown to be effective in influencing specific determinants of behavior were extracted from several studies [59,61,142-144] for each of the phase-specific determinants targeted in the intervention (Table 1). For example, the BCT information about health consequences can be applied to influence the determinants knowledge, risk perception, outcome expectancies, and attitude toward the behavior [142,143]. In addition, it is worth noting that many of the selected BCTs were found to be effective in promoting health behaviors [139,145-149] and may be associated with improvements in HbA1c among people with T2D [146]. For instance, research found that mainly the combination of action planning and coping planning techniques was effective in improving physical activity levels [147]. In another study, physical activity interventions that included the following BCTs showed larger effect sizes at follow-up (ie, maximum of 6 months) than interventions that did not: action planning, instruction on how to perform the behavior, and prompts and cues [148].

Phase 2: Development Process

Overview
We specified the main intervention features for E-Supporter, consisting of (1) goal-setting options (ie, step goals, cycling goals, or nutritional goals) and (2) intervention options consisting of motivational messages, feedback, and reinforcement or barrier identification combined with psychological exercises (Figure 1). To provide insight into and feedback on current lifestyle behavior, the E-Supporter content can be used in combination with self-monitoring tools (eg, Fitbit activity tracker [150], self-reported activities, and digital food diary) of the apps in which E-Supporter is integrated. On the basis of these main intervention features and the results of the definition phase, we determined the tailoring variables, wrote the decision rules, and developed content for the intervention options (Table 2).
<table>
<thead>
<tr>
<th>Tailoring variables</th>
<th>Decision rule</th>
<th>Decision point</th>
<th>Intervention options</th>
<th>Example intervention option</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motivational message based on duration of intervention use</td>
<td>IF duration ≤15 days THEN (initiation phase message) OR (phase generic message) ELSE IF duration &gt;15 days and duration ≤30 days THEN (initiation phase message) OR (action phase message) OR (phase generic message) ELSE IF duration &gt;30 days and duration ≤45 days THEN (action phase message) OR (phase generic message) ELSE IF duration &gt;45 days and duration ≤60 days THEN (action phase message) OR (maintenance phase message) OR (phase generic message) ELSE IF duration &gt;60 days THEN (maintenance phase message) OR (phase generic message)</td>
<td>Twice a day</td>
<td>Informative, motivational, or advisory motivational message</td>
<td>Action phase, disease-generic, morning-specific, and physical activity goal: “Good morning [name], did you know that your environment can encourage you to be more physically active? For example, put your walking shoes by the door. Then you will be reminded to walk!”</td>
</tr>
<tr>
<td>Motivational message based on type of chronic disease</td>
<td>IF type of illness=dna THEN (diabetes-specific message) OR (disease-generic message) ELSE IF type of illness=other THEN (generic message)</td>
<td>Twice a day</td>
<td>Informative, motivational, or advisory motivational message</td>
<td>Initiation phase, diabetes-specific, any time of the day, and physical activity goal: “Hi [name], Did you know that physical activity reduces the risk of additional physical consequences of diabetes (complications)? Examples of these complications include damage to the feet, nerves, eyes, kidneys and heart and blood vessels.”</td>
</tr>
<tr>
<td>Motivational message based on time of day</td>
<td>IF time ≥9 AM and time ≤12 PM THEN (morning-specific message) OR (all moments message) ELSE IF time &gt;12 PM and time ≤6 PM THEN (afternoon specific message) OR (all moments message) ELSE IF time &gt;6 PM and time ≤9 PM THEN (evening specific message) OR (all moments message)</td>
<td>Twice a day</td>
<td>Informative, motivational, or advisory motivational message</td>
<td>Initiation phase, morning-specific, diabetes-specific, and nutrition goal: “Eating breakfast is particularly important. People with diabetes who skip breakfast are on average heavier and have higher blood sugar levels. Your body produces less insulin if you do not eat breakfast.”</td>
</tr>
<tr>
<td>Motivational message based on type of behavioral goal</td>
<td>IF behavioral goal=(physical activity) THEN (physical activity message) OR (all goal message) ELSE IF behavioral goal=(nutrition) THEN (nutrition message) OR (all goal message)</td>
<td>Twice a day</td>
<td>Informative, motivational, or advisory motivational message</td>
<td>Maintenance phase, physical activity, disease generic, and any time of the day: “Hi [name], Is it time to expand your goal? For example, do you now walk 20 minutes on Monday and Wednesday? Then try to walk for 30 minutes on those days. Or you can go for an extra walk, for example on Friday.”</td>
</tr>
<tr>
<td>Feedback based on goal achievement (step goal example)</td>
<td>IF total number of steps≥goal set THEN (motivational feedback) ELSE IF total number of steps&lt;goal set THEN (feedback) AND (identify barriers) AND (recommend intervention)</td>
<td>7 days after setting a goal</td>
<td>Provision of feedback on goal achievement and examination of barriers and recommendation of an exercise</td>
<td>Goal not achieved: “Hi [name], let us look back at how it went last week. You achieved your goal in 3 out of 7 days last week. It seems that you are currently finding it difficult to reach your goal. Why do you find it difficult to achieve your goal?”</td>
</tr>
<tr>
<td>Exercise based on identified barrier</td>
<td>IF identified barrier=motivation THEN (motivation exercise) ELSE IF identified barrier=competence THEN (self-efficacy exercise) ELSE IF identified barrier=mood THEN (mood exercise) ELSE IF identified barrier=stress THEN (stress exercise) ELSE IF identified barrier=planning THEN (planning exercise) ELSE THEN (provide nothing)</td>
<td>7 days after setting a goal</td>
<td>Provision of matching psychological exercise</td>
<td>Mood: “By reflecting on pleasant things that you experience, you become happier. Think of 3 fun things that happened to you today or yesterday. You experienced 3 things that make you happier. If you want, you can make such a list at the end of every day for the next week. Hopefully, this helps you to think positively.”</td>
</tr>
</tbody>
</table>
**Step 1: Tailoring Variables**
In E-Supporter, 6 tailoring variables were applied: duration of intervention use, type of chronic disease, time of day, type of behavior goal, goal achievement, and the identified barrier toward goal achievement.

The first tailoring variable was the duration of intervention use. The duration of use of the intervention was measured by calendar after first login and was used to select an intervention option fitting a specific behavior change phase over time (Figure 3). In addition, intervention options were tailored to the type of chronic disease because people are more likely to follow advice when it is relevant to them [39]. Therefore, intervention options could be tailored to people with T2D or consisted of generic information related to changing health behaviors. This allows tailoring of information to other diseases in subsequent intervention versions. Another tailoring variable comprised the time of day. To better match advice with the time of day, some intervention options contained information appropriate for a particular time of day, namely, in the morning, afternoon, or evening. Other intervention options could be sent at all parts of the day. Time of the day was measured with the smartphone clock. Furthermore, type of behavior goal was used as a tailoring variable. Intervention options were tailored to the type of behavior that an individual wanted to improve (ie, either physical activity or nutrition) based on the type of weekly goal that was set (ie, step goal, cycling goal, or nutritional goal). Goal achievement was also used by tailoring intervention options to the percentage of days per week the behavioral goal was met. Monitoring of goal achievement was based on passive assessment by a Fitbit activity tracker or self-reported goal achievement by means of daily ecological momentary assessments (EMAs) with 24-hour recall. EMA includes repeated sampling of individuals’ current behaviors in real time and in subjects’ natural environments [151]. In addition, goal achievement determined whether individuals were asked about promoting factors or barriers for goal achievement and if an intervention option should be delivered (Figure 1). Finally, the identified barrier toward goal achievement was used to tailor intervention options and was assessed weekly by EMA.

**Step 2: Decision Rules**
The decision rules were operationalized to provide the right type of intervention option tailored to the user circumstances. The decision rules were based on IF-THEN rules specifying the situation (IF) with the cutoff point of a given situation (eg, if a goal is reached or not) and the characteristics of an intervention option (THEN). There were three types of decision rules: (1) rules that triggered the type of motivational message, (2) rules that triggered feedback on goal achievement, and (3) rules that triggered a type of psychological exercise. Table 2 shows examples of decision rules for each intervention option.

**Step 3: Intervention Options**
Motivational Messages
Motivational messages were designed for one-way communication, delivered as push notifications, and written in the Dutch language at the Common European Framework of Reference for Languages B1-level [152]. We developed a database of 425 motivational messages, consisting of content for each of the tailoring variables. Decision points took place at 2 semirandom times per day for a period of 10 weeks.

To be able to tailor motivational messages to the duration of intervention use, a set of motivational messages was written for each phase of behavior change. Message content was based on the determinants and corresponding BCTs identified in the definition phase (Table 1). Each BCT was operationalized using the definitions and examples provided in the Behavior Change Technique Taxonomy, version 1. Examples of motivational messages for each BCT can be found in Multimedia Appendix 3. Addressing the different phases of behavior change over time was reflected as follows: (1) earlier messages focused more explicitly on persuasive messages (targeting attitudes) or awareness raising (targeting knowledge and risk perceptions), (2) later messages focused more on performing new behavior (targeting action and coping planning), and (3) the latest messages focused on behavioral maintenance (eg, targeting habit formation). In addition, the key determinants were addressed throughout the whole duration of the intervention. For the tailoring variable type of chronic disease, content of the messages was divided into advice that applied to everyone (ie, generic messages) and advice that only applied to people with T2D (ie, diabetes-specific messages). Motivational messages that aimed at a specific moment of the day contained information appropriate to that time of day. For example, messages about breakfast were sent in the morning or about taking a lunch walk in the afternoon. To tailor messages to the type of behavior goal, we developed 3 types of messages: goal independent messages, physical activity messages, and nutritional messages. The content of the messages was aligned with the intervention objectives. For example, motivational messages for physical activity mainly focused on promoting light to moderate physical activities that were considered as most feasible for the target group, such as gardening, brisk walking, and cycling [109]. For this, health information was used from books, reliable websites (eg, the website of the Dutch Nutrition Center), national lifestyle guidelines, and diabetes specialists.

During the focus groups with health care professionals, 74.6% (208/279) of messages were directly approved. Furthermore, 20.1% (56/279) of motivational messages were revised by reformulating texts and were approved afterward. In addition, 5.4% (15/279) of messages were excluded from the database. The main reasons for adaptation or exclusion were that messages (1) were too difficult to understand, (2) contained information that only health care professionals are allowed to give, and (3) raised unrealistic expectations of the effects of a healthy lifestyle.

**Feedback**
Feedback based on goal achievement was provided weekly after the last goal assessment moment of the week. Feedback was given regarding whether they achieved their goal, consisting of both descriptive and evaluative feedback. Everyone received feedback on how many days the goal was achieved in the past week. If individuals realized their goal substantially or even their full goal, they received a compliment (eg, “You’re doing well!”). In addition, the users were prompted about promoting
factors by asking “Think about what helped you to work on and achieve your goal this week” and “Is this something which might help you next week as well?” (Figure 1). If individuals had limited achievement of their goal, they received feedback such as “At the moment it seems difficult to achieve your goal. What is the main reason for this?” Thereafter, motivation, self-efficacy, mood, stress, or planning problems were assessed to identify barriers to goal achievement.

**Psychological Exercises**

When a goal was not sufficiently reached and motivation, self-efficacy, mood, stress, or planning was the identified barrier for goal achievement, an appropriate psychological exercise (fitting the indicated barrier) option was selected to support problem-solving. If participants realized their weekly goal sufficiently, they were offered an exercise of choice (eg, self-efficacy), after being stimulated to think about promoting factors that helped them in their goal progress. In both cases, individuals were able to decline the exercise and received good luck wishes for the coming week. Whenever individuals agreed to complete an exercise, a random exercise concerning the chosen determinant was initiated. In total, there were 5 distinct categories of behavioral determinants with varying amounts of related exercises: motivation (n=3), self-efficacy (n=5), planning (n=4), mood (n=8), and stress (n=2). An overview of exercises and related content per determinant can be found in Multimedia Appendix 4.

Exercises comprised a dialog between the user and conversational agent using motivational interviewing techniques (Table 1; BCT 3.1), which is a direct, person-centered conversation style that promotes behavioral change by strengthening an individual’s intrinsic motivation and commitment to change [153]. During a dialog, the user was prompted to think, plan, or elaborate on setting future steps toward the health behavior and reply to questions of the coach by completing open input fields or choosing a predefined answer fitting the user’s response. The response of the coach was selected from a set of possible predefined answers depending on the user’s input. For example, in the exercise importance ruler related to the determinant motivation, users were asked to indicate how important the health behavior was for them and why they think it was important to them. At the end of the exercise, the user received a summary highlighting why engaging in the respective health behavior was important to them. Finally, all users, independent of goal achievement, received the possibility to adapt their weekly goal. Goal setting was guided by the coach and consisted of three options, including (1) preserve the current goal, (2) adapt the current goal (ie, increasing or decreasing the difficulty of goal), or (3) setting a new goal (eg, from step goal to nutritional goal).

**Step 4: Integration of the E-Supporter 1.0**

The content of E-Supporter 1.0 was (partly) integrated into two apps: (1) the Diameter [89,90] and (2) MiGuide [91]. Both apps aim to monitor and coach people with T2D and encourage the adoption of a healthy lifestyle.

**Integration Within the Diameter App**

Diameter is a Dutch app for people with T2D who aim to improve their glucose regulation through lifestyle changes. To date, the Diameter app has been used in research as a blended care intervention in secondary hospital care. All monitoring and coaching components of E-Supporter were integrated into Diameter app. Motivational messages were integrated as push notifications that could be closed by liking, disliking, or the closing button. Goal setting and weekly exercises that focused on barrier identification and problem-solving were integrated in the form of an interactive dialog with a conversational agent (Figure 4).

---

**Figure 4.** Screenshots from the Diameter app. A: home screen with overview of goal achievement; B: example of motivational message as push notification; C: example of start screen of an exercise.
In addition to be coached via the E-Supporter content, individuals using the Diameter app could monitor physical activity (i.e., with an activity tracker and manually), nutrition (i.e., with an electronic food diary), and real-time glucose levels (i.e., by using a Freestyle Libre 2 glucose sensor) [93]. The data collected with the app and sensors were fed back to health care providers so that these data could be used to conduct a more personal and patient-centered conversation based on objective data. Figure 4 shows some screenshots of the Diameter app.

Integration Within MiGuide
MiGuide is an app for people with T2D with the aim to improve lifestyle. The MiGuide app is available in both the Google Play Store and AppStore, is available in Dutch, and is offered in a blended care setting in primary care. Only the weekly psychological exercises aimed at barrier identification and problem-solving were integrated from E-Supporter 1.0 into the MiGuide app. MiGuide uses its own goal-setting options, self-monitoring tools, and short messages similar to E-Supporter. In the app, previously set goals, physical activity (i.e., using an activity tracker and manually), nutrition (i.e., using an electronic food diary), and glucose levels (i.e., entering measured values manually) could be monitored. Coaching was offered through short messages developed by MiGuide and the psychological exercises from E-Supporter. The MiGuide app could be linked to different General Practitioner Information Systems (in Dutch: Huisarts Informatie Systeem). This allowed data to be exchanged between the MiGuide app and the Huisartsen Informatie Systeem. Patients could view their medical file via the app, and health care providers could gain insights into the measurements of their patients so that more personalized care could be offered. Figure 5 shows some screenshots of the MiGuide app.

Figure 5. Screenshots from the MiGuide app. A: activity screen within MiGuide; B: examples of motivational messages within MiGuide; C: example of an exercise to increase motivation.

Phase 3: Review Phase

Participant Characteristics
In total, 9 participants (n=7, 78% male) were included in the usability study. Participants were on average 65.2 (SD 8.7) years old and obese (mean BMI 31.7, SD 3.29 kg/m$^2$). On average, the participants had been diagnosed with T2D for 17 years. Diabetes-related complications were present in 67% (6/9) of participants. In total, 44% (4/9) of participants had used an app to track their physical activity or diet previously at the time of inclusion in the study.

Intervention Use
Self-monitoring Tools of the Diameter App
All (9/9, 100%) participants scanned the Freestyle Libre more than the requested 3 times per day with an average of 11 (SD 2.7) scans per day. Data loss occurred in 11% (1/9) of participants because the Freestyle Libre detached prematurely (after a week) from the upper arm; 11% (1/9) of participants experienced problems with synchronizing the Fitbit with the app. The participants without synchronization problems had an average of data loss of 24.8% (range 0%-66.4%). In addition, 67% (6/9) of participants logged activities in addition to using Fitbit. The food diary was completed by 89% (8/9) of participants for the requested 6 days.

E-Supporter Content
Of the 9 participants, 7 (78%) participants experienced problems receiving the 2 daily motivational messages; 1 (11%) participant received no messages at all, which could be explained by a human error made by the researcher. The other 6 participants received the messages very irregularly (e.g., 1 instead of 2 messages per day or no messages at all for a day). The log data
showed that the push notifications for the messages were sent from the app but that participants mentioned not to receive a notification on their phone. This issue was probably caused by phone settings or notification blockers (ie, push notifications marked as spam are killed by the system). Of the motivational messages of which participants did receive a push notification, 97.1% (298/307) were read. Participants liked 73.8% (220/298) and disliked 12.1% (36/298) of the content of the motivational messages. The content of the remaining messages was perceived as neutral (42/298, 14.1%). No major differences were found between participants. In total, 43% (17/40) of the psychological exercises were completed, with a broad variety in completion rates between participants (range 0%-100%). Furthermore, it was noted in the log data that missing data (eg, missing Fitbit data and missed EMAs) were included in the weekly feedback by stating that a goal was not achieved on the days when there were missing data.

Acceptability

Overview

Experiences with E-Supporter 1.0 integrated within the Diameter app were reduced to two major themes: (1) Content of the E-Supporter 1.0 and (2) Way of delivery via the Diameter. The theme Content of the E-Supporter 1.0 included perspectives on the content of E-Supporter 1.0 (eg, opinions on information and advice provided). The theme Way of delivery via the Diameter was related to the way the E-Supporter functionalities were integrated in the Diameter app and experiences with the additional self-monitoring tools that were offered with the Diameter app.

Content of E-Supporter 1.0

Participants experienced the goal-setting functionalities as motivating because it gave them a concrete purpose to work on. No areas for improvement were identified for the goal-setting functions. Motivational messages were mostly positively rated. Participants called the coaching messages fun, motivating, and informative and believed that it also contributed to lifestyle improvements. The content of the intervention options matched their preferences regarding the type of physical activities and diet well, but the participants felt that the content could be further tailored. Preferably, the participants would also like to receive real-time feedback on their actual behavior (eg, feedback whether a certain amount of physical activity is sufficient). Opinions were divided about the psychological exercises (ie, conversational agent). Some participants saw the added value of these exercises and thought it made them think about how to achieve their goals as participant 6 mentioned the following:

*I like the online coach. Then you become more aware of your own behavior. I mainly use it to think about how I want to take more steps and the coach does make you think about that.* [Male, 52 years]

However, other participants found the purpose of the weekly exercises unclear which demotivated them to complete the exercises. Participant 3 explained why the conversational agent was not of added value for her:

*I think that is a bit of a nagging of “the coach wants to talk to you”* [push notification to complete the psychological exercise]. I do not need that. Then I have that message again and then I think quickly write in some answers and then we are done with it. I don’t see the added value in that, because I can also think for myself why I will or will not achieve my goal. [Female, 54 years]

Way of Delivery via the Diameter App

The participants considered the self-monitoring functionalities of the Diameter app to be valuable because they provided new insights into their own lifestyle. For example, several participants indicated that through self-reporting and tracking physical activities they learned that they were not as active as they thought. In addition, all the participants indicated that the self-monitoring functionalities gave them insight into the effects of lifestyle on glucose levels as participant 2 said the following:

*The Diameter provides insight, for example which activities you have undertaken, which food you have eaten and what impact that has had on glucose values. That is extremely useful.* [Male, 66 years]

These kinds of insights convinced the participants that an improvement in lifestyle can lead to improved glucose regulation. For several reasons, participants expected that the Diameter app could be a valuable addition to regular care. First, participants stated that the Diameter app could support the transition to a healthy lifestyle by complementing the information and advice of health care professionals during the consultations in the hospital. Second, participants thought that it could be beneficial if health care providers could also access the collected data. Participants believed that insight into these data could allow health care professionals to provide more personalized lifestyle advice. Furthermore, some participants noted that sharing data with health care providers could also lead to more proactive care as participant 6 outlined the following:

*If healthcare providers have insight into my data, they can act much more proactively. This means that I do not have a consultation with the doctor every few months, but that consultations will be planned, when necessary, for example if my blood sugars are poorly regulated. But also, regular consultations will be omitted if everything goes well.* [Male, 52 years]

Participants indicated that they experienced a high degree of user-friendliness because each component (ie, physical activity, nutrition, and glucose levels) had its own tab and there were a limited number of buttons. The biggest point of criticism regarding ease of use was filling in the food diary. It was difficult to find certain foods or these were not available at all in the food diary. This was particularly experienced by participants who often eat dishes from foreign cuisine. As a result, participants had to look for alternative foods that are similar. This took a lot of time, and participants felt that this gave a distorted picture of their diet. Other frequently mentioned disadvantages were related to the way E-Supporter 1.0 was integrated within the Diameter app. It was not possible to read the motivational messages again once they had closed the message as participant 8 echoed the following:
Sometimes I wanted to read the information from the messages again later or I wanted to look up the hyperlink to a website again. That was not possible now. That is a pity because then I cannot do anything with it anymore. [Female, 73 years]

Regarding the weekly psychological exercises, participants found it inconvenient that the exercises came at a fixed day and time in the week and could not be postponed to another moment. In addition, the exercises popped up automatically on their screen, making it impossible to perform another action within the app (e.g., filling in the food diary) without completing the exercise. Participants would like to be able to choose at what time they perform the exercises so that they could also take the time to go through it carefully.

All detected bugs, inconveniences, and so on were listed and fixed by the app developers accordingly whenever possible.

Discussion

Principal Findings

This study describes the development of E-Supporter 1.0, a lifestyle monitoring and coaching intervention, using a systematic and participatory 3-phase design approach. The aim of E-Supporter 1.0 was to encourage people with T2D to be physically active and to follow national dietary guidelines. The HAPA model and theories explaining behavioral maintenance were used to select determinants of behavior and identified BCTs that were presumed to affect the targeted determinants. Thereafter, the intervention was developed by (1) selecting targets to tailor the intervention, (2) operationalizing decision rules to provide the right type of intervention option to an individual, and (3) creating intervention options to influence health behavior.

Regarding intervention development, we ensured a systematic and participatory design approach. The use of program-planning models provided detailed guidance on how to develop E-Supporter 1.0. This approach increased transparency in the design process by providing a comprehensive description of the intervention rationale and development of intervention components. This contributes to a better interpretation of results and the replicability of the intervention [59] and may serve as inspiration for other researchers [154]. Furthermore, health care professionals and people with T2D participated at several moments in the development process. Several studies [155-158] noted the importance of involving end users and other stakeholders in activities related to the development, implementation, and evaluation of eHealth interventions. Development "with" end users or other stakeholders increases the chances of successful adoption of and engagement with eHealth interventions, which in turn increases the likelihood of achieving desired effects. In our study, the involvement of health care professionals and people with T2D provided useful input regarding the requirements, development, and improvement of the intervention. For example, because of the focus groups with health care professionals, the content, readability, and comprehensibility of the motivational messages were improved so that these may have a better fit with the target group. The overall development approach, using program-planning models and participatory design, can facilitate future adjustments and development of the intervention.

So far, most of the eHealth interventions to promote health behaviors have shown positive short-term effects [40-46]. E-Supporter goes beyond many existing eHealth interventions by integrating 3 evidence-based elements that could increase intervention effectiveness. First, theory-driven methods were used as the fundament for the intervention [46,60-63]. Most eHealth interventions mainly focus on intention forming [67,159]. However, individuals often do not act in accordance with their intentions (i.e., intention-behavior gap) [160], and behavior often cannot be maintained in the long term [112]. Therefore, our intervention also focuses on behavior initiation and maintenance in addition to intention forming [111,112,145,161] by covering determinants (e.g., coping planning) and BCTs that target postintentional phases [65,117]. Second, dynamic tailoring was applied to increase the probability of adherence to and effectiveness of the intervention [69,70]. Dynamically tailored interventions provide support that better meet user needs than static tailored interventions [39]. Therefore, dynamic tailoring may increase feelings of personal relevance and responsiveness to the intervention option. Third, our intervention content was integrated into app-based interventions that are used in a blended care setting in primary and secondary care. There is a growing body of literature that recognizes that blended interventions are more likely to be used and effective than stand-alone interventions [36,75-77,79-81]. By combining aforementioned elements, we expect that our intervention could positively contribute to sustainable health behavior change, although this still needs to be researched.

In addition to individual factors, behavior is largely influenced by the (social) living environment [162-164]. Both the physical (e.g., food supply and availability of sports and recreational facilities) and social environment (i.e., the behavior of people in the environment) can contribute to the formation of certain barriers to a healthy lifestyle [24,162]. E-Supporter offers tools and techniques on how to deal with these (social) environmental barriers, for example, by indicating how individuals can organize social support in daily life or can learn to deal with social influences on lifestyle choices. By using eHealth, people’s attitudes toward unhealthy lifestyle behaviors can change and people can learn to deal with barriers in the social environment through coping strategies. However, in many cases, the (social) living environment will not change substantially (e.g., the presence of sports facilities and social influences). A (social) living environment that tempts unhealthy behavior therefore remains an important barrier to successful adoption and maintenance of a healthy lifestyle.

A feature that distinguishes E-Supporter from other eHealth interventions is that E-Supporter content can be integrated in different eHealth interventions and settings. Although this version aimed to improve physical activity and diet in people with T2D, we expect that the content can be used to promote other lifestyle behaviors in people with other chronic diseases with simple adjustments owing to about 80% of the content consisting of nondisease-specific health information. Moreover, often the same behavioral determinants influence the behavior change process (e.g., self-efficacy).
Strengths and Limitations

The strength of this study is that our intervention combines several potentially effective elements for eHealth interventions, including the application of behavior change theory and dynamic tailoring, the deployment of the intervention in a blended care setting, and early end user involvement in both intervention development and evaluation. Another strong point of E-Supporter is that the content can be built into different apps so that it can be used in different contexts and possibly also on a larger scale.

Some limitations are worth noting at this stage of the research. The weekly feedback on goal achievement and whether to offer psychological exercises relies highly on input from the user. Lifestyle goals other than step goals were actively monitored through daily EMAs, which rely on an individual’s daily response. Therefore, work is being done on this issue by making more use of passive assessment tools that require minimal user input (eg, using activity trackers to track cycling goals). In addition, if individuals provided insufficient input, the Diameter app based the tailored feedback on incomplete information. This led to individuals receiving inappropriate feedback that they had not sufficiently achieved their goals, which can be demotivating to use the intervention. The Diameter app was technically adjusted so that no feedback will be given if there is insufficient user input to provide valid feedback.

We had little influence over the design and user interface of the Diameter app into which the E-Supporter content was integrated. The design and interface of the intervention can influence the user experience and use of the app both positively and negatively [165,166]. Our usability test showed that the interface of certain E-Supporter elements in the Diameter app negatively influenced the experience with E-Supporter (eg, automatic pop-up of the psychological exercises). However, it is not clear to what extent a different design and interface (eg, through integration in another app such as MiGuide) will lead to different findings regarding the acceptability of E-Supporter. The content of E-Supporter remains unchanged, but some aspects of the intervention (eg, attractiveness or ease of use) may be experienced differently.

During the usability study, we encountered some challenges in the technical integration of E-Supporter 1.0 into the Diameter app (eg, receiving motivational messages irregularly). These challenges led to the intervention not being delivered as intended, which negatively influenced the results regarding use and acceptability of E-Supporter. To improve the integration, we recorded all detected technical problems and discussed them with the app developer to solve them. Lessons learned from this usability test will be used to make recommendations regarding the integration of E-Supporter content in apps to promote a positive user experience. For example, it is necessary to test what influence missing data has on the feedback initiated by the app (eg, as described in the first limitation).

Future Research

We have planned several follow-up activities to further improve E-Supporter. First, the intervention content and intervention period will be expanded so that E-Supporter 1.0 can better facilitate behavioral maintenance. Literature states that behavior maintenance is reached when an individual can maintain the desired behavior for at least 6 months [167]. To achieve behavior maintenance, it is challenging to offer digital coaching over a long period without losing adherence to the intervention, given the high attrition rates in eHealth use over time [168-170].

Second, researching the use of BCTs in other delivery modes (eg, videos and voice messages) than textual coaching is suggested. The content of the intervention options can remain unchanged but will be offered via a different delivery mode. Other delivery modes can make the intervention options more attractive to increase acceptance [171] and more comprehensible to people with low (health) literacy.

Third, we intend to tailor our intervention more dynamically by applying data science techniques because higher degrees of tailoring can contribute to improved user engagement and effectiveness [31,63,172,173]; for instance, by tailoring the intervention content to additional determinants of behavior, individual characteristics (eg, health literacy), current behaviors, or predicted high-risk situations (eg, as in just in-time adaptive interventions) [39]. To optimize tailoring strategies in future, we can examine the effects of intervention options on proximal outcomes (ie, short-term goals) per individual and whether these effects vary with time or circumstances using study designs such as microrandomized trials [39,174]. In microrandomized trials, individuals are randomized hundreds of times over the course of the study by being randomly assigned to an intervention option at a decision point (ie, time points when an intervention decision must be made). We are already working in a multidisciplinary research team involving, among others, behavioral health experts, health care professionals, and computer science specialists, which is required to develop highly tailored behavior change interventions [31,39,175].

Fourth, our intervention can be improved by aligning intervention content of the ABCC tool and E-Supporter (Figure 1). E-Supporter can be expanded with content for additional target groups (eg, people with chronic obstructive pulmonary disease or chronic heart failure) and lifestyle behaviors (eg, smoking) that are part of the ABCC tool. Moreover, data exchange between the ABCC tool and E-Supporter should be made possible so that the eManager can be used as an integrated blended care intervention.

To obtain more information about the intervention use and acceptability of our intervention, we plan to evaluate the E-Supporter content in other apps, such as MiGuide [91]. This option allows us to investigate whether another design, interface, and functionality will result in different findings on some aspects (eg, attractiveness or ease of use) regarding the experience with the E-Supporter. At the end of 2022, the effectiveness of E-Supporter 1.0 will be explored in a blended care setting through a single-arm longitudinal study with 6-month follow-up. Finally, new versions of E-Supporter will have to be iteratively evaluated regarding user engagement and cost-effectiveness in the long run [31].
Conclusions
This paper describes the systematic and participatory development of a theory-based, dynamically tailored lifestyle coaching intervention to support physical activity and a healthy diet in people with T2D. Program-planning models and behavior change theory were used complementarily during the development of the intervention. The intervention was evaluated in a small usability study which provided insights into intervention use and acceptability. Future work should focus on improving the degree of tailoring and evaluating its effects on acceptability, use, and cost-effectiveness.

Acknowledgments
We thank the health care providers and patients who participated in the development phase and usability study for their active participation and input. In addition, we thank Anouk ten Voorde and Elise Fokkema for their contribution to writing and improving the motivational messages.

This study was funded by the Netherlands Organization for Health Research and Development (grant 104006001). The funders in the study had no role in the study design, data collection, analysis and interpretation of the data, or writing of the report.

Authors' Contributions
EAGH mostly contributed to drafting the manuscript and worked closely together with AM, PvE, KP, and AAJK during the design, development, and usability testing of the intervention. WON-d’H and LKS contributed significantly to the development of intervention content. GDL and MMRV-H provided ongoing feedback related to the methods, results, and discussion of this paper. All authors critically evaluated and approved the definitive version of the paper.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Methods E-Supporter health expert review.
[PDF File (Adobe PDF File), 118 KB - humanfactors_v10i1e40017_app1.pdf ]

Multimedia Appendix 2
Methods E-Supporter usability study.
[PDF File (Adobe PDF File), 171 KB - humanfactors_v10i1e40017_app2.pdf ]

Multimedia Appendix 3
E-Supporter determinants and behavior change techniques (BCTs) including examples.
[PDF File (Adobe PDF File), 248 KB - humanfactors_v10i1e40017_app3.pdf ]

Multimedia Appendix 4
E-Supporter psychological exercises.
[DOCX File , 26 KB - humanfactors_v10i1e40017_app4.docx ]

References


Abbreviations

ABCC: Assessment of Burden of Chronic Conditions
BCT: behavior change technique
EMA: ecological momentary assessment
HAPA: health action process approach
QoL: quality of life
T2D: type 2 diabetes

©Eclaire A G Hietbrink, Anouk Middelweerd, Pepijn van Empelen, Katharina Preuhs, Annemieke A J Konijnendijk, Wendy Oude Nijeweme-d’Hollosy, Laura K Schrijver, Gozewijn D Laverman, Miriam M R Vollenbroek-Hutten. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 12.01.2023. This is an open-access article distributed under the terms

https://humanfactors.jmir.org/2023/1/e40017 JMIR Hum Factors 2023 | vol. 10 | e40017 | p.34
(page number not for citation purposes)
Exploring the Use of Pictograms in Privacy Agreements to Facilitate Communication Between Users and Data Collecting Entities: Randomized Controlled Trial

Larissa Ugaya Mazza¹, BSc, MSc; Laura Xavier Fadrique¹, BSc, MSc; Amethyst Kuang¹, BSc, MHI; Tania Donovska², MSc; Hélène Vaillancourt³, PhD, PEng; Jennifer Teague², PhD; Victoria A Hailey⁴, BA; Stephen Michell⁵, BMATH, MSc; Plinio Pelegrini Morita¹,²,⁶,⁷, MSc, PhD, PEng

¹School of Public Health Sciences, Faculty of Health, University of Waterloo, Waterloo, ON, Canada
²Institute of Health Policy, Management, and Evaluation, University of Toronto, Toronto, ON, Canada
³Canadian Standards Association Group, Montreal, QC, Canada
⁴Victoria Hailey Group Corporation, Toronto, ON, Canada
⁵Canadian Standards Association Group, Etobicoke, ON, Canada
⁶eHealth Innovation, Techna Institute, University Health Network, Toronto, ON, Canada
⁷Department of Systems Design Engineering, University of Waterloo, Waterloo, ON, Canada

*these authors contributed equally

Corresponding Author:
Larissa Ugaya Mazza, BSc, MSc
School of Public Health Sciences
Faculty of Health
University of Waterloo
Hagey Blvd Toby Jenkins Applied Health Research Building
Waterloo, ON, N2L 0A7
Canada
Phone: 1 226 201 3065
Email: lari.u.ma@gmail.com

Abstract

Background: Privacy agreements can foster trust between users and data collecting entities by reducing the fear of data sharing. Users typically identify concerns with their data privacy settings, but due to the complexity and length of privacy agreements, users opt to quickly consent and agree to the terms without fully understanding them.

Objective: This study explores the use of pictograms as potential elements to assist in improving the transparency and explanation of privacy agreements.

Methods: During the development of the pictograms, the Double Diamond design process was applied for 3 instances of user interactions and 3 iterations of pictograms. The testing was done by performing a comparative study between a control group, which received no pictograms, and an experimental group, which received pictograms. The pictograms were individually tested to assess their efficacy by using an estimated comprehension of information symbols test.

Results: A total of 57 participants were recruited for the pictogram evaluation phase. With the addition of pictograms, the overall understanding improved by 13% (P=.001), and the average time spent answering the questions decreased by 57.33 seconds. A 9% decrease in perceived user frustration was also reported by users, but the difference was not significant (χ²=4.80; P=.31). Additionally, none of the pictograms passed the estimated comprehension of information symbols test, with 7 being discarded immediately and 5 requiring further testing to assess their efficacy.

Conclusions: The addition of pictograms appeared to improve users’ understanding of the privacy agreements, despite the pictograms needing further changes to be more understandable. This proves that with the aid of pictographic images, it is possible to make privacy agreements more accessible, thereby allowing trust and open communication to be fostered between users and data collecting entities.

Trial Registration: ClinicalTrials.gov NCT05631210; https://clinicaltrials.gov/ct2/show/NCT05631210
Introduction

Privacy agreements fulfill the important role of helping users understand how their data will be used by data collecting entities [1]. The role of privacy agreements is to not only provide users with the chance to decide whether they want to disclose their data to an entity but also foster trust and reduce users’ concerns about data sharing [1,2].

Many users are concerned about personal data collection, and privacy agreements may alleviate these concerns. However, due to the complexity of privacy agreements, there are barriers to understanding data use [3], which result in users agreeing to terms that they do not fully comprehend [4]. This paper explores the use of pictograms as a potential way to improve the transparency of privacy agreements and users’ understanding of privacy agreements.

Most studies about pictograms used as communication tools focus on pictograms that depict pharmaceutical- and health-related information [5-19] or hazardous substances and their safe handling [7,20-31].

Pictograms are useful when communicating certain types of information for which language, literacy, and reaction times can be barriers [32]. For example, some studies have shown that pictograms are beneficial for facilitating danger recognition and the understanding of precautionary measures [5,7,8,12,13,18,22,31,33]. Some of the advantages of using pictograms instead of written words are that they can facilitate faster recognition and remembrance during a second encounter and can improve the understanding of communicated messages for people with visual deficiencies or low literacy levels and people who are unfamiliar with the language used. Chief among these advantages is that pictograms can be more easily understood than their written counterparts [8,9,18,20,26,28,30,31]. When it comes to health care, pictograms have been shown to be better at informing patients about examination preparation [34].

Nevertheless, pictograms are not the solution for all communication problems. As Spinillo [35] argues, pictograms should be used judiciously, since images are more appropriate for representing material things, relative sizes, and simultaneous concepts. However, they are often inadequate for representing general or abstract concepts [35,36].

In this paper, pictograms will be defined as “graphic images that immediately show the user of a hazardous product what type of hazard is present. With a quick glance, [the user] can see, for example, that the product is flammable, or if it might be a health hazard” [37]. Pictograms are composed of both graphic and textual parts. The graphic parts include the border and the symbol, that is, a black image inside the border [37]. The textual part comprises bolded text indicating the name of the pictogram and a legend (in brackets) with a description of the hazard.

Because images are not a global language, they cannot be used for different population groups without the risk of losing or changing their meaning [8,10,20,32,35]. This makes it important to consider the specific target group when developing pictograms and to rigorously test pictograms throughout the design process [5,6,9,10,13,19,20,28,31,35].

When considering pictograms overall, Wogalter [28] talks about the four main purposes of a warning in his book *Handbook of Warnings*. In it, he says that a warning must (1) communicate important safety information; (2) influence or modify a person’s behavior to improve their safety; (3) reduce or prevent accidents, injuries, damage, or health problems; and (4) serve as a reminder for those that are already aware of the danger.

There are 4 components in the warning context that affect a pictogram’s creation and implementation [26], as follows: (1) the source (the designer, sender, or originator of the warning message), (2) the medium (how the message is being displayed; eg, visual, auditory, etc), (3) the message (the content), and (4) the receiver (the target audience that the warning seeks to reach).

According to Laughery and Wogalter [26], for a message to flow effectively, it must go from one component to the next in a linear fashion. If the connection is severed at any point, the flow can be broken, resulting in the failure to deliver the warning.

To avoid this, the involvement of users is imperative not only for testing but also for the design process. User involvement is invaluable for the inclusion of previously overlooked elements that result in improved performance [5,6,8-10,12,14,16,19,28,31,32].

A warning can have many different parts, and it is the role of the designer to combine them effectively [26]. Each component may serve different purposes and change when directed at different users. An example of this would be using more technical language when dealing with specialists but using simpler terminologies with novices [26].

With regard to the visual components, a semiotic study separated them into the following two categories: transparent and opaque components [20]. Transparent symbols highly resemble their real-life counterparts, have guessable meanings, are useful when communicating internationally, and are more easily understood than abstract images [9,15,21,32,33]. However, they cannot accurately represent abstract concepts, such as emotions or situations, which are dependent on cultural contexts.

Opaque symbols on the other hand do not have a clear relationship with their referents [20]. Although they can represent complex and abstract concepts, they may not be immediately recognizable and must be learned beforehand [31,32,35].
Symbols can also be separated into the monosemic, polysemic, and pansemic image categories [6]. Monosemic images only have 1 meaning, polysemic images have 2 or more meanings, and pansemic images have many meanings. Typically, abstract images are pansemic; however, monosemic connotations are preferable when developing a pictogram to communicate information about hazardous substances [6].

Pictograms based on existing systems have more transparent connotations [35,38]. Thus, as a person becomes more familiar with a certain type of visual language, they become more apt at interpreting different pictograms, provided that the pictograms follow the same visual synthesis [35,38].

In this study, requirements were taken from the designs of hazard and health-related pictograms for the development of the pictograms that were used to facilitate privacy agreement understanding. The requirements are as follows: (1) making the pictograms with users; (2) testing the pictograms with users; (3) developing the pictograms by using an iterative method; (4) building upon pictograms from existing systems; (5) using pictures with labels, keywords, or short texts; (6) using color; and (7) making the pictograms culturally relevant.

The objectives of this study were to develop a set of pictograms that represent the top 10 privacy concerns, assess their impact on users’ understanding, and encourage users to engage with privacy agreement content. The hypothesis is that with the incorporation of visual assistance, the users will find reading privacy agreements easier and less frustrating.

Methods

Design Method for the Development of Pictograms

This research was part of a larger project that focused on trust and privacy agreements. The larger project was divided into the following three phases: (1) identifying the top 10 privacy concerns, (2) exploring the use of pictograms for privacy agreements, and (3) assessing the effectiveness of the new privacy agreement layout. This research focused on phases 2 and 3, using the results gathered from phase 1.

The methodology that was used to develop the pictograms was based on the Double Diamond design methodology (Figure 1). It was chosen for its iterative nature and the many points of contact between the designers and users.

The methods were divided into the following four phases: the Discover, Define, Develop, and Deliver phases. During the Discover phase, the research scope was expanded to understand the users’ needs and opinions. The Define phase was used to narrow the scope and analyze the collected data to identify trends, themes, and patterns. In the Develop phase, techniques such as brainstorming, sketching, and graphic recording were used to further develop previously identified ideas. The scope was closed a final time in the Deliver phase, during which a solution was prototyped and tested by users.

In this study, contact with users only took place during the Discover, Develop, and Deliver phases. A total of 9 participants were included in a visualization exercise for the Discover phase. They were asked to sketch their ideas for the visual representations of the top 10 privacy concerns on paper and to briefly explain what they were thinking when they made these representations. These visualizations were analyzed for trends and patterns during the Define phase. Afterward, based on these patterns, the pictograms were constructed during the Develop phase. Lastly, in the Deliver phase, the pictograms and a version of a privacy agreement that implemented them were validated by a group of users through a questionnaire.

![Double Diamond design method](image)

Evaluation of Overall Understanding and User Frustration

An evaluation was conducted to test whether the addition of the pictograms made reading privacy agreements more efficient and less frustrating for users. For this purpose, a questionnaire was developed along with 2 versions of the privacy agreement. The control group (31 participants) received the traditional version of the privacy agreement while the experimental group received the version of the privacy agreement that included the pictograms (29 participants).

The survey was closed and distributed through Amazon Mechanical Turk—a website that allows people to fill out surveys for a small monetary gain. The administration of the survey was performed via Amazon Mechanical Turk, and security for the survey and the assurance that there were no duplicate responses were provided by the website. All questions were multiple-choice questions, and if there was a question that was not properly filled, the data for that whole entry were discarded, which happened only once.

In total, 62 people started the survey and 57 people completed it. The target population was people who had some understanding of technology, and the sample was a convenience
The data were collected during the first week of September 2019.

Both groups were quizzed on the content of their version of the privacy agreement and were later asked to rate their perceived level of frustration when looking for the answers. Participants were then asked for suggestions about changes to the privacy agreement and the pictograms.

The 4-part questionnaire was developed by using Qualtrics (Qualtrics International Inc)—a web-based tool—and beta tested via a pilot study to assess its feasibility. The first part asked demographic questions about participants’ age, sex, ethnicity, occupation, education, country of residence, and region. We used the second part to compare the performance of the control group to that of the experimental group for part 3. In the second part, the control group was given the traditional version of the privacy agreement, whereas the experimental group was given the version of the privacy agreement with a group of pictograms that summarized its content, which appeared before the written section. Both versions of the privacy agreement can be viewed in Multimedia Appendix 1.

Participants were then asked to answer 5 questions that quizzed them on the content of the privacy agreement that they had received. For both groups, all questions were about the information represented by the pictograms.

The questions were as follows, and each question was given its own page on the survey:

- **Question 1:** “Is your information being collected?”
- **Question 2:** “Can you opt out of some services?”
- **Question 3:** “Will your data be identifiable when shared?”
- **Question 4:** “Is your location being collected?”

- **Question 5:** “Can third parties have access to your data?”

Each participant’s response was timed to assess how quickly participants could find the correct answers based on the information presented in their version of the privacy agreement. Time data were compared between the control group and the intervention group.

The third part of the questionnaire asked participants to rate their frustration levels while answering part 2, their level of concern, and their previous knowledge about data privacy. In total, there were 9 pages in the survey, which included the option to return to the previous pages before the end of the survey.

**Evaluation of Pictogram Efficiency**

In the fourth part of the survey, participants were asked to take an estimated comprehension of symbols test [39] to measure how comprehensive the pictograms were for public use and to determine what further revisions would be required.

Each pictogram was presented individually, coupled with a description of what it was supposed to represent, without a legend. Participants were asked to rate the percentage of the population that they thought would be able to understand the pictogram and the description.

For cases with an estimated comprehension level of <47%, the symbol was considered a failure. On the other hand, an estimated comprehension level of >87% was deemed appropriate. An estimated comprehension level of between 47% and 87% implied false negatives and false positives. For such cases, the symbol would have to be tested again by using a classic comprehension test [39]. An example of how the fourth part of the questionnaire looked can be found in Figure 2.

**Ethics Approval**

The survey was cleared by the University of Waterloo ethics board (application number: 4060 Privacy Agreement for Sharing Health Data), and was registered with Clinical number NCT05631210. The survey was voluntary, and participants could stop participating at any moment. At the start of the questionnaire, the participants were told about the purpose of the study, its length, the possible risks, and the benefits of taking the survey. They were then asked for informed consent. The only personalized information collected was employment status, sex, age, ethnicity, and the places where participants lived.
Results

Participants
A total of 57 participants were recruited by using Amazon Mechanical Turk; 28 completed the questionnaire with the privacy agreement that implemented the pictograms, and 29 completed the one with the original, imageless privacy agreement.

The sample consisted of 22 female participants and 35 male participants who resided in the United States (n=18), Canada (n=19), or Europe (n=18). The Europeans were from the United Kingdom (n=6), the Netherlands (n=2), Italy (n=3), Germany (n=2), France (n=3), Spain (n=1), and Estonia (n=1). The distribution of ethnicities was White (n=49), Black (n=3), Chinese (n=2), South Asian (n=1), Southeast Asian (n=1), and Filipino (n=1). The rest of the participants’ demographics and occupations are described in Multimedia Appendices 2 and 3.

Different levels of interest in data privacy were reported; 21 participants reported high levels of concern about data privacy, 20 expressed moderate concern, 10 had low concerns, and 6 were neutral. Further, 35 participants thought that data privacy was highly important, 13 considered it to be moderately important, 4 believed it had little importance, and 5 were neutral.

With regard to previous knowledge about data privacy, 3 participants claimed to be highly knowledgeable, 29 claimed to have moderate knowledge, 14 claimed that they had little knowledge, 8 felt neutral about their knowledge, and 3 claimed to have no knowledge. These variations in concerns and knowledge levels can be reviewed in Figure 3.

Figure 3. Concerns and knowledge about data privacy reported by participants.

Development of Pictograms

Overview of Pictogram Development
A representation exercise was conducted during a workshop with 9 members of the Ubiquitous Health Technology Lab at the University of Waterloo. The participants were informed of the top 10 privacy concerns one at a time and were asked to create visual representations that they felt would accurately embody each concern.

In total, 90 pages of visualizations with varying degrees of representational content were collected to represent 10 privacy concerns. Figure 4 shows an example of the representations collected for one of the privacy concerns—“Is my location being collected?” Each white page belonged to a single participant, and the colored stickers represented the most common elements across the sample. Finally, the blue papers summarized the most used elements within the sample for a given privacy concern.

A content analysis, which followed the Define phase of the Double Diamond design method, was completed to organize and identify the patterns and trends within the representation ideas. The most used elements throughout the visualizations were (1) arrows or the notion of direction (n=56), (2) representations of the self (n=41), (3) clouds (n=23), and (4) binary code (n=17).

The first set of pictograms was developed by using the results from the visualization exercise, and they can be found in Multimedia Appendix 4. The pictograms were based on material design icons, in accordance with one of the guidelines sourced from the literature [40]. There were 3 iterations of pictograms, and with each iteration, implementation feedback was solicited from the team of design professionals.

We used one-on-one interviews to acquire feedback, during which the context was explained to the participants. They were given a single sheet of paper with all of the pictograms printed on it. After the interviews, the researcher asked each participant to explain what they thought each pictogram represented. Afterward, the researcher told the participant what the intended meanings were and asked them to propose changes that they believed would improve comprehension.

The responses were audio-recorded and then analyzed to detect whether the participants had guessed the meaning of a pictogram correctly. After the first interview, a second set of pictograms was developed, and this can be found in Multimedia Appendix
5. This set passed through the same interview process as the previous set to create the final set (Figure 5).

The final set of pictograms was divided into the following three categories: pictograms that addressed what the user could choose to do, pictograms that presented facts that could not be changed by the user, and pictograms that showed the user what the system permitted them to do.

**Figure 4.** Representations collected during the workshop to represent the concept of “is my location being collected?”

![Figure 4](image)

**Figure 5.** Final set of pictograms developed.

---

**User Possibility: What the User Can Choose**

The first category of pictograms showed the user what options were available for them to choose. In this category, the pictograms were for “data being collected,” “data will be deleted after the deletion of the app/account,” “location is being collected,” “data collected is anonymized,” “data being sold,” and “data can be shared with third parties.”

These pictograms were designed differently from those in the other two categories. They had a yellow frame shaped like a square on one of its axes to mimic warning pictograms. This shape was chosen to attract more attention, since these pictograms showed the user what information they had control over.

---

**User Impossibility: What the User Cannot Choose**

The second category contained pictograms that presented characteristics of the system that the user had no control over. The pictograms in this category were counterparts to all of the pictograms in the `User Possibility` category, with additional pictograms for “microphone will have access to your data,” “camera will have access to your data,” and “your data will be collected for academic purposes.” These pictograms had the same core black and white symbol but had a circular blue frame.

---

**System Characteristics: What the System Lets the User Do**

This category of pictograms showed what the system allowed the user to do. The pictograms in this category were “opt-out,”
“data access through your computer,” and “data access through your phone.” These pictograms had the same blue circular frame as those in the User Impossibility category.

**Overall Understanding of the Privacy Agreements With Pictograms**

A 2-tailed Pearson correlation test was performed to assess if there was a relationship between privacy concerns and knowledge. There was a slightly positive correlation between privacy concerns and knowledge, but it was not significant ($r=0.10, \text{df}=98; P=.87$), as shown in Figure 6.

Introducing pictograms improved the overall understanding of privacy agreements by 13%. The original layout resulted in 106 right answers, 28 wrong answers, and 11 people who did not know the answer. The layout with the pictograms resulted in 121 right answers, 8 wrong answers, 3 people who did not want to read the privacy agreement, and 8 people who did not know the answer. Participants in the experimental group chose the correct answer 13% more often than the control group but chose the “didn’t know the answer” option 2% less often than the control group. However, they also chose the “didn’t want to read” option 2% of the time, while no one selected the same option in the control group.

Fisher exact tests (Table 1) were performed for each answer type to determine differences in levels of understanding between the two groups. This study found that participants’ understanding was significantly associated with the privacy agreement layout with which they were presented ($P=.001$).

Although the increased accuracy of answers that was observed with the addition of the pictograms was not significant ($P=.008$), this improvement demonstrates that participants still had a better understanding of the privacy agreement content if images were presented alongside text.

![Figure 6. Relationship between privacy concerns and knowledge.](image)

<table>
<thead>
<tr>
<th>Answer type</th>
<th>Understanding of privacy agreement with pictograms, SE</th>
<th>Understanding of original privacy agreement, SE</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right answer</td>
<td>0.03</td>
<td>0.03</td>
<td>.008</td>
</tr>
<tr>
<td>Wrong answer</td>
<td>0.01</td>
<td>0.02</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Did not want to read privacy agreement</td>
<td>0.01</td>
<td>0</td>
<td>.12</td>
</tr>
<tr>
<td>Did not know answer</td>
<td>0.01</td>
<td>0.01</td>
<td>.64</td>
</tr>
</tbody>
</table>

**Time Spent Reading the Privacy Agreements**

The amount of time spent reading the privacy agreement and answering the five questions decreased for all questions except for the first one. Moreover, 1-way Mann-Whitney U tests (Table 2) were conducted for each question to investigate whether the decreases in time between the two privacy agreement versions were significant. The only significant decreases in time were observed for questions 2 ($P<.001$) and 4 ($P=.004$).

The average time for answering all 5 questions decreased by 57.33 seconds with the addition of the pictograms, suggesting that their addition assisted users in finding the correct answers faster. A potential reason for this could be that users did not read the original privacy agreement and only used the summarized information that the pictograms displayed, or the pictograms helped users understand the presented information better. However, considering that the overall understanding of the privacy agreement improved with the addition of the pictograms, it is more likely that they helped users gain insight.
into privacy agreements while also decreasing the time spent searching for specific information.

Table 2. The 1-way Mann-Whitney U test for the time spent reading the privacy agreement.

<table>
<thead>
<tr>
<th>Question</th>
<th>Privacy agreement with pictograms</th>
<th>Original privacy agreement</th>
<th>Wilcoxon test statistic</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question 1</td>
<td>Time spent (seconds), median 48.66</td>
<td>38.17</td>
<td>481.00</td>
<td>.89</td>
</tr>
<tr>
<td>Question 1</td>
<td>Time spent (seconds), mean (SD; SE) 87.40 (112.13; 21.19)</td>
<td>51.53 (51.41; 9.55)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question 2</td>
<td>Time spent (seconds), median 8.80</td>
<td>39.44</td>
<td>193.00</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Question 2</td>
<td>Time spent (seconds), mean (SD; SE) 18.08 (23.14; 4.37)</td>
<td>61.48 (65.59; 12.18)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question 3</td>
<td>Time spent (seconds), median 8.99</td>
<td>30.90</td>
<td>330.00</td>
<td>.12</td>
</tr>
<tr>
<td>Question 3</td>
<td>Time spent (seconds), mean (SD; SE) 22.81 (34.49; 6.52)</td>
<td>51.94 (84.35; 15.66)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question 4</td>
<td>Time spent (seconds), median 8.40</td>
<td>16.78</td>
<td>239.00</td>
<td>.004</td>
</tr>
<tr>
<td>Question 4</td>
<td>Time spent (seconds), mean (SD; SE) 11.34 (11.70; 2.21)</td>
<td>31.52 (42.87; 7.96)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question 5</td>
<td>Time spent (seconds), median 6.58</td>
<td>5.46</td>
<td>366.00</td>
<td>.27</td>
</tr>
<tr>
<td>Question 5</td>
<td>Time spent (seconds), mean (SD; SE) 12.71 (24.84; 4.69)</td>
<td>13.19 (14.64; 2.72)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Perceived Frustration While Reading the Privacy Agreements**

Users in the experimental group reported experiencing less frustration compared to the control group. There was 9% less perceived frustration in the experimental group. For the original layout, 24 people were neutral in terms of frustration, 23 reported being a little frustrated, 15 were very frustrated, and 3 were extremely frustrated. For the layout with the pictograms, 31 participants were neutral, 18 were a little frustrated, 13 reported being frustrated, 16 were very frustrated, and 6 were extremely frustrated. Average levels of frustration (“a little bit frustrated” and “frustrated”) decreased by 14% with the addition of the pictograms. However, high levels of frustration (“very frustrated” and “extremely frustrated”) increased by 5% in the experimental group when compared to those in the control group.

A chi-square test was performed to investigate if there were significant differences in perceived frustration levels between the two groups, and a 2-tailed Pearson correlation test was performed to assess if there was a relationship between the frustration and privacy concern levels. The chi-square test (Table 3) showed that overall frustration levels were not significantly different between the two layouts ($\chi^2=4.80$; $P=.31$), and the 2-tailed Pearson correlation (Figure 7) test showed that there was a slight negative correlation between privacy concerns and frustration levels for the original version of the privacy agreement, though the negative correlation was not significant ($r=-0.05$, df: 98; $P=.80$).

Table 3. Chi-square test ($\chi^2=4.80$; $P=.31$) and Fischer exact test for frustration.

<table>
<thead>
<tr>
<th>Frustration level</th>
<th>Frustration with privacy agreement with pictograms, SE</th>
<th>Frustration with original privacy agreement, SE</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutral</td>
<td>0.03</td>
<td>0.03</td>
<td>.25</td>
</tr>
<tr>
<td>A little bit frustrated</td>
<td>0.02</td>
<td>0.03</td>
<td>.48</td>
</tr>
<tr>
<td>Frustrated</td>
<td>0.02</td>
<td>0.03</td>
<td>.13</td>
</tr>
<tr>
<td>Very frustrated</td>
<td>0.02</td>
<td>0.02</td>
<td>.84</td>
</tr>
<tr>
<td>Extremely frustrated</td>
<td>0.01</td>
<td>0.01</td>
<td>.32</td>
</tr>
</tbody>
</table>
Figure 7. Relationship between privacy concerns and frustration (pictograms).

Estimated Comprehension of Information Symbols

None of the pictograms passed the estimated comprehension of information symbols test; 7 pictograms were discarded, as they had a score of less than 47%, and the remaining 5 pictograms were scored between 47% and 87% by participants and required further validation via a comprehension test (Table 2). The pictogram with the highest rating was “microphone is accessing your data,” with a 62.8% level of estimated comprehension, and the pictogram with the lowest rating was “your data is being collected.” A summary of the scores for all pictograms can be found in Multimedia Appendix 6.

Pictograms that relied on established material design icons and used transparent symbols based on suggestions made by Berthenet et al [9], Vaillancourt et al [16], Mok et al [33], Spinillo [35], and Mayer and Law [21] for designing pictograms had the best reception.

The ones that scored lower were the pictograms that used opaque symbols with pansemic meanings, which make pictograms harder to understand before they are incorporated into common knowledge [6,9,16,21,33]. A summary of which pictograms scored less than 47% and which ones scored between 47% and 87% can be found in Multimedia Appendix 7.

Discussion

This research aimed to explore the use of pictograms for privacy agreements and assess the effectiveness of the new privacy agreement layout. Our findings suggest that the addition of pictograms improved the users’ experiences with understanding a privacy agreement when searching for information, even when suboptimal pictograms were provided. The decrease in the time taken to find the correct information and the self-reported decreased levels of frustration and confusion when engaging with the privacy agreements suggest a positive correlation between the addition of pictograms to privacy agreements and the perceived transparency of the documents’ contents.

To summarize, using images as an explanatory tool may improve the overall user experience when reading a privacy agreement and may even increase the understanding of the information being presented.

Even though the users considered none of the pictograms to be highly intuitive, the addition of the pictograms still helped users find the information about their data privacy settings, even when the pictograms’ meanings were less than transparent. We can assume that with the passage of time, these symbols will become integrated into common knowledge, will facilitate more interest in reading privacy agreements, and will result in such documents becoming more accessible to the general public, thereby fostering both trust and communication between users and the entities that collect their data.

Acknowledgments

This project was developed alongside the CSA Group and received funding from them.

Conflicts of Interest

None declared.

Editorial notice: This randomized study was only retrospectively registered, as the authors believed registration was unnecessary. The editor granted an exception from ICMJE rules mandating prospective registration of randomized trials, because the risk of bias appears low. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related
to primary outcomes or effectiveness, as retrospective registration does not prevent authors from changing their outcome measures retrospectively.

Multimedia Appendix 1
Control group privacy agreement versus experimental group privacy agreement.

Multimedia Appendix 2
Demographics of participants.

Multimedia Appendix 3
Occupation of participants.

Multimedia Appendix 4
First set of pictograms developed.

Multimedia Appendix 5
Second set of pictograms developed.

Multimedia Appendix 6
Estimated comprehension of pictograms.

Multimedia Appendix 7
Pictograms to be discarded or further validated according to the estimated comprehension test.

References
6. da Silva CRL, Spinillo C, Soares M. As contribuições da linguagem gráfica pictórica para o design de avisos e advertências em medicamentos. 2007 Presented at: Congresso Internacional de Pesquisa em Design; October 11-13, 2007; Rio de Janeiro, Brazil. [doi: 10.13140/2.1.3699.5528]
7. Ta GC, Mokhtar MB, Mokhtar HABM, Ismail AB, Yazid HFBBHA. Analysis of the comprehensibility of chemical hazard communication tools at the industrial workplace. Ind Health 2010;48(6):835-844 [FREE Full text] [doi: 10.2486/indhealth.ms1153] [Medline: 20616463]


37. Canadian Centre for Occupational Health and Safety. URL: https://www.ccohs.ca/ [accessed 2019-08-04]


Edited by A Kushniruk; submitted 24.02.22; peer-reviewed by F Wirth, M Jungkunz; comments to author 23.04.22; revised version received 03.06.22; accepted 04.06.22; published 25.01.23.

Please cite as:
Ugaya Mazza L, Fadrique LX, Kuang A, Donovska T, Vaillancourt H, Teague J, Hailey VA, Michell S, Morita PP
Exploring the Use of Pictograms in Privacy Agreements to Facilitate Communication Between Users and Data Collecting Entities: Randomized Controlled Trial
JMIR Hum Factors 2023;10:e34855
URL: https://humanfactors.jmir.org/2023/1/e34855
doi:10.2196/34855
PMID:36696167

©Larissa Ugaya Mazza, Laura Xavier Fadrique, Amethyst Kuang, Tania Donovska, Hélène Vaillancourt, Jennifer Teague, Victoria A Hailey, Stephen Michell, Plinio Pelegrini Morita. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 25.01.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Development of a Digital Behavioral Intervention to Reduce the Consumption of Sugar-Sweetened Beverages Among Rural Appalachian Adults: Multiphased, Human-Centered Design Approach

Jamie Zoellner¹, RDN, PhD; Annie Reid¹, RDN, MPH; Kathleen Porter¹, RDN, PhD; Christina Frederick², MA; Michelle Hilgart², PhD; Lee Ritterband², PhD

¹Department of Public Health Sciences, School of Medicine, University of Virginia, Christiansburg, VA, United States
²Department of Psychiatry and Neurobehavioral Sciences, School of Medicine, University of Virginia, Charlottesville, VA, United States

Corresponding Author:
Jamie Zoellner, RDN, PhD
Department of Public Health Sciences
School of Medicine
University of Virginia
16 East Main Street, Suite 101
Christiansburg, VA, 24073
United States
Phone: 1 4349624488
Email: jz9q@virginia.edu

Abstract

Background: To avoid the low engagement and limited efficacy of digital behavioral health interventions, robust human-centered design (HCD) processes are needed.

Objective: The primary objective of this study was to describe a flexible, step-by-step HCD process to develop digital behavioral health interventions by illustrating iSIPsmarter as an example. iSIPsmarter is a digital intervention for reducing the consumption of sugar-sweetened beverages (SSBs) that comprises 6 internet-based cores metered out over time to deliver the program content, an integrated SMS text message strategy to engage users in reporting SSB behaviors, and an electronic cellular-enabled scale for in-home weighing. The secondary objective is to illustrate the key components and characteristics of iSIPsmarter that resulted from the HCD process.

Methods: The methods were guided by the Model for Internet Interventions and by best practices in HCD and instructional design processes (eg, rapid prototype development and think-aloud protocol). The 3-phased (ie, contextual, prototype testing, end user testing phases) process followed in this study included a series of 13 semistructured one-on-one interviews with 7 advisory team participants from the targeted Appalachian user group. The interviews were content coded by 2 researchers and then deductively coded to the suggested areas of digital behavioral health interventions.

Results: The participants provided rich perspectives pertaining to iSIPsmarter’s appearance, behavioral prescriptions, burdens, content, delivery, message, participation, and assessment. These inputs included requests for built-in flexibility to account for varying internet and SMS text message accessibility among users; ideas to resolve the issues and problems encountered when using the prototypes, including those related to navigation and comprehension of content; ideas to enhance personalized feedback to support motivation and goal setting for SSB consumption and weight; and feedback to refine the development of realistic and relatable vignettes. The participants were able to interact with multiple prototype drafts, allowing researchers to capture and incorporate feedback related to the iSIPsmarter dashboard, daily SSB and weight diaries, action planning, core content, interactions, and vignettes.

Conclusions: Using scientific models and established processes is critical for building robust and efficacious interventions. By applying an existing model and HCD and instructional design processes, we were able to identify assumptions and address the key areas of the iSIPsmarter intervention that were hypothesized to support users’ engagement and promote behavior change. As evidenced by the rich feedback received from the advisory team members and the resulting iSIPsmarter product, the HCD methodology was instrumental in the development process. Although the final iSIPsmarter content is specific to improving SSB consumption among rural Appalachian adults, the HCD process and its outcomes are broadly applicable to the development of digital behavioral health interventions.
consumption behaviors among adults in rural areas, the intent is that this HCD process will have wide applications in the development of digital behavioral health interventions across multiple geographic and behavioral contexts.

**KEYWORDS**
eHealth; human-centered design; internet-based intervention; digital technology; Model for Internet Interventions; beverages; behavioral research; rural population; mobile phone

**Introduction**

**Background**
The availability of digital behavioral health interventions has surged in recent years. However, when deployed in research trials and disseminated in real-world practice, the uptake of and engagement with many digital interventions are often lower than desired [1-3]. In turn, this results in the unrealized potential of both immediate and sustained health outcome improvements among the intended users [3]. Although numerous factors may contribute to low uptake, one of the most important is a suboptimal fit between the characteristics of the technology and the needs, skills, and context of the user. To adequately address these complex interrelationships, the development of digital behavioral health interventions should include a theory-driven, iterative, and human-centered design (HCD) process.

Importantly, there are a number of models and frameworks available to guide the digital intervention development processes [4-9]. The *Model for Internet Interventions* [4] is a key example that has been used as a basis for the development of many digital health programs [10-13]. This model can help researchers distinguish and operationalize various components of digital behavioral health interventions and identify the relationships among the components. More specifically, this model posits that to explain behavior change across digital behavioral health interventions, it is necessary to consider design-related components, areas, and elements, including user characteristics, environment, intervention content, level of intervention support, and targeted outcomes [4]. Furthermore, the *Model for Internet Interventions* highlights 8 main areas that comprise the digital health application used to deliver the intervention (ie, appearance, behavioral prescriptions, burdens, content, delivery, message, participation, and assessment). To improve the likelihood of digital health applications meeting the needs and requirements of users, HCD and instructional design processes should be applied when considering and manipulating these 8 areas.

Although this and other models and frameworks provide helpful principles and guidelines [4-9], they are not intended to be a step-by-step prescription for digital intervention development. A scoping review of 160 papers regarding research activities for participatory eHealth development processes identified a variety of methods and products [14]. However, there was no evidence of an optimal single-step approach for developing digital behavioral health intervention applications [14]. The findings from this narrative review highlighted the importance of researchers and developers selecting the most appropriate objectives for and methods for developing the context and user characteristics of their digital behavioral health interventions.

Despite the flexibility in methodological processes, the application of HCD and instructional design processes in the development of digital behavioral health interventions is consistently recommended [15]. The International Organization for Standardization defines HCD as “an approach to systems designs and development that aims to make interactive systems more usable by focusing on the use of the system and applying human factors and usability knowledge and techniques” [16]. Interdependent design activities include understanding and specifying the context of use, specifying the user requirements, producing design solutions, and evaluating the design. In addition, the International Organization for Standardization posits that six requirements must be met for an HCD process: (1) the design is based upon an explicit understanding of users, tasks, and environments; (2) users are involved throughout design and development; (3) the design is driven and refined by user-centered evaluation; (4) the process is iterative; (5) the design addresses the entire user experience; and (6) the design team includes multidisciplinary skills and perspectives [16].

Complementary to HCD concepts, instructional design processes involve designers executing cycles of continuous formative evaluation to ensure that the intervention meets the users’ needs, prior knowledge, and experience [15]. These instructional design processes should involve setting measurable learning objectives or performance requirements, assessing the users’ achievement of the targeted outcomes, and revising the program components until the desired outcomes are achieved [15,17-20]. Ultimately, the ability of digital behavioral health interventions to achieve sustained engagement and desired behavioral outcomes is enhanced when using scientific frameworks along with proven and context-specific HCD and instructional design approaches during the development process.

**Appalachian Digital Environment and Sugar-Sweetened Beverage Context**

Although HCD processes are important across different contexts, the need for them is magnified in rural regions, including Appalachia. Historically, extending digital interventions into the Appalachian region has been hindered by digital divide [21,22]. However, similar to other rural American communities, Appalachia is making great strides in narrowing the digital divide and closing the rural-urban gap in home broadband internet connection and smartphone ownership [23-26]. However, little is known about how rural Appalachian adults engage with digital behavioral health interventions. As such, human-centered and instructional design processes are especially important for interventions targeting this and similarly underserved regions where access to digital behavioral health interventions has been limited.
In addition to the geographic and digital divide context of Appalachia, the behavioral context of sugar-sweetened beverage (SSB; eg, soda or pop, sweet tea, sports and energy drinks, and fruit drinks) consumption is noteworthy. SSBs are the single largest source of calories in the US diet and account for approximately 7% of the total energy intake of the US adults [27]. In Appalachia, SSB intake is disproportionately higher and accounts for an average of 14% of the total energy intake—twice as high as national estimates [28]. Consistent with both national and Appalachia-specific data showing that SSB is the largest contributor to added sugar intake [27-29], excessive SSB consumption is undeniably pervasive in Appalachia. However, there are only a few known SSB consumption–specific behavioral interventions that have used digital technologies or applied HCD processes [30-33].

**Objectives**

The primary objective of this study was to describe a flexible, step-by-step approach to and an HCD process for developing digital behavioral health interventions by illustrating iSIPsmarter as an example. iSIPsmarter is a technology-based behavior and health literacy intervention aimed at improving SSB consumption behaviors among Appalachian adults. The secondary objective is to illustrate the key components and characteristics of iSIPsmarter that resulted from the HCD process. Although iSIPsmarter’s content is specific to improving SSB consumption behaviors among Appalachian adults, the intent is that the HCD process will have wide applications in the development of digital behavioral health interventions across multiple geographic and behavioral contexts.

**Methods**

**Overview**

The development process for the digital behavioral health intervention iSIPsmarter was guided by the Model for Internet Interventions [4] and by best practices in HCD [16] and instructional design processes [17-20], which is being evaluated in a randomized controlled trial (Trial Registration: NCT05030753). Figure 1 illustrates the conceptual integration of this model and these processes, along with key definitions. Specifically, the development process of iSIPsmarter included 13 semi-structured interviews across three non-sequential iterative phases: (1) contextual, (2) prototype testing, and (3) end user testing phases.

This formative, flexible, and step-by-step approach allowed the advisory team participants to provide insights at all phases of the development process and interact with multiple prototype drafts developed by the content development team (CDT).

The following sections describe the considerations for the CDT, advisory team participants, adaptation context, iSIPsmarter intervention, and adaptation process. Subsequently, the 3 phases of data collection and the data analysis strategy are described.

**Ethical Considerations**

This program evaluation project involved activities that do not represent human participant research and, therefore, did not require submission to the institutional review board (Human Research Protection Program Standard Operating Procedures [34]).

**The Role of CDT**

The development of digital behavioral health interventions should be conducted by a multidisciplinary CDT [15,16,35]. This includes subject matter experts, behavioral psychologists, web designers, instructional designers, and developers and programmers.

The multidisciplinary CDT involved in the development of iSIPsmarter included 4 doctoral-level researchers: 2 (50%) SSB content experts who had led prior SSB behavioral intervention trials in the targeted Appalachian region, 1 (25%) expert on digital behavioral health technology interventions, and 1 (25%) instructional design expert. This team was further supported by research staff experienced in nutrition, public health, and digital 

---

**Figure 1.** Conceptual overview and integration of the Model for Internet Interventions and best practices in human-centered design processes.
health development and implementation, including those with expertise in the design and creation of robust user interface and user experience. For the duration of this HCD process, the team met weekly to collaboratively draft advisory team interview guides with clear objectives, develop prototypes, and respond to feedback from the participants. Likewise, the CDT also wrote content for the iSIPsmarter cores, developed media-rich interactions and videos, drafted personas and vignettes, and developed an integrated internet-based platform.

Advisory Team Participants

When developing digital behavioral health interventions, the selection of advisory team members is an important decision. Ideally, the participants should represent the intended end user of the intervention in terms of demographic and cultural characteristics, health literacy and digital literacy skills, geographic location, and patterns of the targeted behaviors [16].

The advisory team members involved in the development and design of iSIPsmarter were a convenience sample of prior SIPsmartER participants who resided in the Appalachia region. The advisory team members also had to have internet access and the ability to review the iSIPsmarter program materials on the web. These members were diverse in terms of age (21 to 60 years), gender, and socioeconomic status. Of the 15 prior SIPsmartER participants invited, 7 (47%) joined the advisory team. These participants were involved throughout the intervention design and development processes, with the first interview conducted in February 2019 and the final interview in December 2021. In total, the advisory team members were compensated with up to US $200 in electronic gift cards for their participation.

Adaptation Context

iSIPsmarter is a technology-based behavior and health literacy intervention aimed at improving SSB behaviors among Appalachian adults [36]. iSIPsmarter was adapted from the evidence-based intervention SIPsmartER [37]. SIPsmartER is a 6-month intervention that includes 3 in-person small-group classes, 1 live-teach-back call, and 11 interactive voice response (IVR) calls; the intervention also comprises a 12-month maintenance phase that includes monthly IVR calls (refer to the database Evidence-Based Cancer Control Programs for details [38]). The classes cover key behavioral content and action planning, whereas the IVR calls engage users in reporting SSB intake and personal action planning. Several formative research projects among Appalachian adults initially guided the development of SIPsmartER [39-41]. The content and strategies of SIPsmartER are guided by the theory of planned behavior and health literacy, numeracy, and media literacy concepts [42-46]. In brief, SIPsmartER was found to be effective at reducing and maintaining SSB consumption behaviors through a full-scale randomized controlled trial and a pilot dissemination and implementation trial conducted in collaboration with the Virginia Department of Health districts in Appalachia, Virginia [29,37,47-58]. Although SIPsmartER had been highly successful in reducing SSB consumption, transitioning from its original structure to a fully digital modality presented an opportunity to focus on optimizing scalability and reach.

At the onset of developing iSIPsmarter, the CDT had ample experience understanding and intervening on SSB consumption in the targeted region and had over a decade’s worth of rich qualitative, observational, and experimental data from the previous SIPsmartER trials. As such, content and behavior change techniques related to SSB consumption were relatively well established when embarking on the digital intervention development. However, little was known about the opportunities, barriers, and access to technology in the targeted population for intervention delivery or data collection purposes. Similarly, weight self-monitoring and the use of cellular-enabled scales to encourage in-home weighing had not been previously explored in the targeted population and region. Therefore, the intention of the iSIPsmarter adaptation process was largely focused on changing the mode of delivery and adding key content around weight while maintaining other core components and the cultural relevance for the intended Appalachia target audience.

iSIPsmarter Description

In its final form, iSIPsmarter comprises 6 internet-delivered interactive cores, which are metered out sequentially over time, with a new core becoming available 1 week after completion of the previous core; an integrated SMS text message strategy to engage users in tracking and reporting SSB consumption behaviors; and an electronic cellular-enabled scale for in-home weighing [36]. At the end of cores 2 to 6, the users evaluate their SSB and weight diary data and set a personalized SSB consumption and weight action plan for the upcoming week. After the sixth core is completed, there is a recurring maintenance core where users can continue behavioral tracking and personalized action planning. To assist with mastering key content and behavioral strategies, the cores also include PDF resources that users can view or print. iSIPsmarter is highly interactive and contains a media-rich format of text, audio, graphics, animation, interaction, and video. Vignettes are woven throughout the intervention, which include stories (based on the experiences of past SIPsmartER participants) to model and describe situations related to setting goals, behavior changes, resolving barriers, and encountering slips. Finally, iSIPsmarter includes a stepped care engagement strategy (ie, human-supported text messages, followed by phone calls if needed) to provide encouragement, technical assistance, and strategies to promote core completion. The adaptation processes, HCD approach, and findings leading to the final content and structure of iSIPsmarter are further detailed in the subsequent sections.

Areas of Adaptation and the Digital Health Intervention Development Process

The Model for Internet Interventions highlights 8 areas to consider when developing digital behavioral health interventions (Figure 1) [4]. These areas are defined and illustrated within the adaptation context for iSIPsmarter.

Content and Messages

Content refers to the actual intervention information and may be the single most important component of the program [4]. Focus is placed on the style and source of messages, which are
theorized to impact user engagement and other mechanisms of change, including knowledge acquisition and motivation. [4]. Similar to SIPsmartER, iSIPsmarter is grounded in the Theory of Planned Behavior and health literacy, numeracy, and media literacy concepts. In addition, the scientifically grounded core content, key learning objectives, and behavior change techniques of the digital iSIPsmarter intervention are similar to those of the original evidence-based SIPsmartER intervention [36,37]. The main difference between the two is the addition of evidence-based weight self-monitoring and weight-related strategies to iSIPsmarter. Therefore, the development process focused heavily on weight-related content and messages. In addition, all iSIPsmarter content was written using clear communication strategies and with the goal of achieving eighth grade reading level. Specifically, all user-facing core content pages were entered into the Readable website to determine the readability score using the validated Flesch-Kincaid Grade Level formula, which considers sentence length and word length [59]. iSIPsmarter passages were revised in up to 2 rounds to improve readability (ie, reduce sentence length and reduce multisyllabic words). This process resulted in 90% of the iSIPsmarter core content pages meeting the criteria of eighth grade level.

**Delivery and Participation**

Delivery refers to the form in which the intervention content is disseminated and includes text, audio, illustrations or graphics, animations, video, and vignettes or stories or testimonials [4]. Participation is focused on the program’s ability to engage and involve the user in the intervention [4]. Within each iSIPsmarter core, videos and interactions were developed using gamification principles to encourage the users to engage in learning and practicing key behavioral strategies and techniques [60]. The intervention content prioritized both interactions focused on skill development and personalized reinforcement to drive behavior change (refer to Table 1 for details). To replicate the role modeling and observational learning that occurred during the SIPsmartER small-group classes, personas and vignettes were developed for iSIPsmarter. Personas are defined as user archetypes that represent the characteristics of future users or actual people from a targeted group [14]. Vignettes are narrative stories that illustrate key situations and real-life scenarios faced by individuals and problem behaviors that the intervention aims to improve. The development of personas and vignettes was a multistep and iterative process. First, based on the CDT’s experience with past SIPsmartER participants and intended users of the digital iSIPsmarter intervention, 9 personas were identified and developed to represent the range of traits and SSB consumption change patterns that the CDT witnessed among SIPsmartER participants. These personas detailed numerous user characteristics, such as (1) demographics and family or social characteristics, (2) SSB consumption change patterns and barriers, (3) weight-related patterns and barriers, (4) motivation level and perceived behavioral control, and (5) the use of planned digital technology components. Then, journey maps were created to share with the advisory team participants. These journey maps were single-page snapshots modeling each persona’s potential path through iSIPsmarter, including experiences, successes, and challenges with improving SSB consumption and weight behaviors. Feedback was solicited from the advisory team participants. On the basis of the participants’ insights, the personas were narrowed down and further refined. Next, vignettes (ie, narrative stories) were developed for each persona to model and describe situations related to setting goals, behavior changes, resolving barriers, and encountering slips. Finally, vignettes were mapped to key content through each of the iSIPsmarter cores and several of the interactions.
Table 1. Final iSIPsmarter overview: asset summary, user objectives, interactions or video descriptions, printable PDFs, and user summaries.

<table>
<thead>
<tr>
<th>Core</th>
<th>Overview and number of assets</th>
<th>User objectives</th>
<th>Interactions or videos description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core 1: getting ready</td>
<td>Core content screens: 26</td>
<td>See how iSIPsmarter works and what to expect</td>
<td>Interactive questions to raise awareness on SSB(^a) availability, SSB costs, and the amount of sugar in SSBs</td>
</tr>
<tr>
<td></td>
<td>Action planning screens: N/A</td>
<td>List my personal reasons for joining iSIPsmarter</td>
<td>Interactive content to show successful results from the previous SIPsmartER Intervention</td>
</tr>
<tr>
<td></td>
<td>Interactions: 4</td>
<td>Discover what counts as a sugary drink</td>
<td>Sorting game to recognize and practice what counts as an SSB</td>
</tr>
<tr>
<td></td>
<td>Videos: 3</td>
<td>Recognize portion sizes of sugary drinks</td>
<td>Sorting game to recognize the portion size of SSBs</td>
</tr>
<tr>
<td></td>
<td>Vignettes: 16</td>
<td>Recall my typical sugary drink patterns</td>
<td>Three short videos to highlight the importance of tracking SSB consumption, the importance of tracking weight, and how to use iSIPsmarter to track SSB consumption and weight</td>
</tr>
<tr>
<td></td>
<td>Printable documents: 3</td>
<td>Track my sugary drinks and weight for my iSIPsmarter Diaries</td>
<td></td>
</tr>
<tr>
<td>Core 2: making a plan</td>
<td>Core content screens: 29</td>
<td>View the recommendations for sugary drinks</td>
<td>Interactivity to rate the importance of and confidence in decreasing SSB consumption, with personalized feedback</td>
</tr>
<tr>
<td></td>
<td>Action planning screens: 22</td>
<td>Recognize the health risks of too many sugary drinks</td>
<td>Interactive game to realize the amount of sugar in SSBs and equivalent sugar packets per day</td>
</tr>
<tr>
<td></td>
<td>Interactions: 4</td>
<td>Explore red-light, yellow-light, and green-light drink categories</td>
<td>Interactive body map to recognize the health risks and key health facts associated with the consumption of too many SSBs</td>
</tr>
<tr>
<td></td>
<td>Videos: 1</td>
<td>See the health benefits of non-sugary drinks</td>
<td>Video to learn about how the consumption of too many SSBs impacts the body and leads to health risks over time</td>
</tr>
<tr>
<td></td>
<td>Vignettes: 12</td>
<td>Evaluate my weight and see a healthy weight range</td>
<td>Interactivity to illustrate the connection between the consumption of SSBs and weight over time</td>
</tr>
<tr>
<td></td>
<td>Printable documents: 8</td>
<td>Set a personal Action Plan to help meet my sugary drink and weight goals</td>
<td></td>
</tr>
<tr>
<td>Core 3: using numbers</td>
<td>Core content screens: 23</td>
<td>Recognize my calorie and energy balance needs</td>
<td>Video to highlight the key components and application of the nutrition facts label, which includes grams of added sugars, servings per container, serving size, and ingredient list</td>
</tr>
<tr>
<td></td>
<td>Action planning screens: 8-12</td>
<td>Identify my limits on added sugars</td>
<td>Four-part interactivity to apply one's skills in reading nutrition labels, identifying different names for sugars, and sorting drinks using the nutrition facts label and to learn tips that can be applied when a drink does not have a nutrition facts label</td>
</tr>
<tr>
<td></td>
<td>Interactions: 1</td>
<td>Apply information from food labels to identify sugary drinks</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Videos: 1</td>
<td>Set a personal Action Plan to help meet my sugary drink and weight goals</td>
<td>Two-part interactivity to identify red-light foods that are consumed often and select red-light foods that can be removed, replaced, or reduced</td>
</tr>
<tr>
<td></td>
<td>Vignettes: 9</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Printable PDFs: 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Core 4: balancing choices</td>
<td>Core content screens: 33</td>
<td>Explore red-light, yellow-light and green-light food categories</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Action planning screens: 8-12</td>
<td>See how red-light drinks and foods can create imbalance in my diet</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Interactions: 1</td>
<td>Recognize how to plan ahead to balance my choices</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Videos: 0</td>
<td>Practice reducing my red-light drink and food choices</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vignettes: 9</td>
<td>Discover the health benefits of physical activity</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Printable documents: 6</td>
<td>Set a personal Action Plan to help meet my sugary drink and weight goals</td>
<td></td>
</tr>
</tbody>
</table>
### Behavioral Prescriptions and Assessments

Behavioral prescriptions instruct the user on how to address the targeted problem. They are designed to increase commitment and boost adherence and may include, for example, behavioral contracts as well as automated and personalized prompts (eg, emails and SMS text messages) [4]. Assessment refers to the ability to measure the needs of the user, personalize the program, and provide tailored content and recommendations [4]. In the case of iSIPsmarter, the personalized action planning process, behavioral monitoring, and personalized feedback loop were transitioned from the original format of small-group classes and IVR calls in SIPsmarterER to a fully digital format. The web development team built iSIPsmarter on a proprietary software platform called the Research Infrastructure Containing eHealth (RICE) interventions, developed for building digital health programs. The RICE platform integrates all aspects of a digital intervention for a seamless user experience and research administration. iSIPsmarter provides an opportunity to extend the technological capability of the intervention infrastructure by incorporating SMS text message and sensor integration, specifically, the BodyTrace scale (BodyTrace, Inc). Daily SMS text message prompts are sent to encourage users to report the of ounces of SSBs consumed the previous day. In addition, an electronic cellular-enabled BodyTrace scale is provided for weight data collection. Users are encouraged to step on the scale daily, with a minimum threshold of 3 days per week, to receive personalized feedback. The integration of the internet-based platform, SMS text messages, and cellular-enabled scales created an integrated experience, whereby users can log into the internet-based platform and view their progress through the cores, along with synced personal diary data (SSB consumption and weight) on their iSIPsmarter dashboard. When appropriate, gamification principles are also applied when providing tailored feedback and recommendations (eg, cues when cores are complete and encouraging feedback when SSB and weight diaries are entered) [60].

### Burdens

Burdens are specific to the intervention content and can include problems related to use, such as poor application navigation and program length [4]. In iSIPsmarter, understanding the burden of the users was prioritized when developing features associated with diary tracking, dashboard navigation, the action plan process. Given that goal setting, planning, self-monitoring, and feedback are among the most important behavior change techniques, understanding and resolving potential user burdens with the digital content was especially important. In addition, attention was paid to the overall length of each core and to balancing the amount of content with the projected user fatigue.

### Appearance

Appearance refers to the look and feel of the application and can include the use of color, page or screen layout, organization of content, and screen size [4]. For iSIPsmarter, a graphic designer developed key images and icons with a consistent theme and style. These graphics were applied throughout the intervention, including all core screens and printable documents.

### Additional Components of the Digital Health Intervention Development Process

The *Model for Internet Interventions* (Figure 1) also illustrates other design-related components and elements (eg, support, user characteristics, environment, and digital health intervention use) [4]. Many of these components were considered in the iSIPsmarter development process. As an example, the model describes *supports* (eg, email, phone, and face-to-face) that can directly impact users’ adherence to digital health applications.

---

**Table: Core Content and User Objectives**

<table>
<thead>
<tr>
<th>Core</th>
<th>Overview and number of assets</th>
<th>User objectives</th>
<th>Interactions or videos description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core 5: thinking critically</td>
<td>Core content screens: 26  Action planning screens: 8-12  Interactions: 2  Videos: 0  Vignettes: 10  Printable documents: 2</td>
<td>Recognize my sugary drink choices may be swayed by marketing  Identify common marketing tactics used to sell sugary drinks  Discover the goal of sugary drink companies is to make me spend money  Evaluate the costs of sugary drinks  Analyze, critique, and modify sugary drink ads  Set a personal Action Plan to help meet my sugary drink and weight goals</td>
<td>Interactivity to recognize that, in the context of SSB marketing, people’s attention is the product  SSB commercial with interactive questions and feedback to help the participants see through marketing and think critically about advertisements</td>
</tr>
<tr>
<td>Core 6: staying on track</td>
<td>Core content screens: 23  Action planning screens: 8-12  Interactions: 1  Videos: 0  Vignettes: 12  Printable documents: 2</td>
<td>Evaluate my sugary drink and weight program goals  Learn how to handle slips in my red-light sugary drink and food choices  Focus on strategies to prevent prolonged relapses  Check my understanding of key iSIPsmarter concepts  Set a personal Action Plan to help me stay on track with my changes</td>
<td>A 25-item interactive quiz to help the participants review and check their understanding of the key iSIPsmarter concepts in cores 2 to 6</td>
</tr>
</tbody>
</table>

---

\(^{a}\text{N/A: not applicable.}\)

\(^{b}\text{SSB: sugar-sweetened beverage.}\)
During the development and design of iSIPsmarter, interviews were conducted with the advisory team members to assess their understanding and perceptions of program reminders and gather inputs on the timing and content of stepped care messages.

Data Collection

Overview

The iSIPsmarter components were developed using a flexible and iterative 3-phased HCD process. This process included semistructured interviews with the advisory team participants. For each interview, the CDT collaboratively developed key objectives and designed interview guides during weekly team development meetings. One of the researchers drafted the initial questions based on the focus area and agreed upon objectives. Subsequently, other team members provided feedback until a final version was agreed upon. All the interview guides used open-ended questions and probes. The interview content was prioritized by the current development activities. Therefore, it was possible to follow an ongoing, iterative, HCD process that matched the steps and pace of the program development. One or 2 researchers led the audio-recorded interviews. The interviews were completed over telephone and via Zoom (Zoom Video Communications Inc). The sessions ranged from 45 to 60 minutes.

Contextual Phase

The contextual phase of the HCD process involves gathering information from the intended users regarding their behaviors and their context and requirements for using the technologies through which the digital health interventions will be implemented [5,16,18]. Given the CDT’s ample experience intervening in SSB consumption among frequent consumers in the targeted region and a deep understanding of many of the components of the Model for Internet Interventions (ie, user characteristics, mechanisms of change, and behavior change) [4], the 4 interviews comprising the contextual phase largely focused on important technology-related factors associated with potential iSIPsmarter use as well as research evaluation components.

The initial advisory team meeting explored participants’ technology ownership and use, internet availability and use, and perceived benefits of and barriers to receiving a digital behavioral health intervention. The 3 subsequent contextual interviews assessed the participants’ understanding and perception of program reminders and gathered feedback on the timing and content of stepped care messages. Likewise, the participants’ experiences with daily weighing on scales and perception toward using cellular-enabled scales for a digital health intervention were evaluated. Finally, the participants’ feedback on the research outcome data components that were assessed was solicited via an internet-based survey, telephone calls, and cellular-enabled scales, all of which were completed by the participants in the comfort of their homes.

Prototype Testing Phase

Prototype testing includes the visual representation of the to-be-developed technology, where the CDT explored different concepts and possible solutions with the intended users. Testable prototypes can take many forms, such as paper prototypes, mock-ups, and wireframes (eg, a skeletal framework of an interface, usually a website or other applications) [5,16,18]. Prototype testing is a critical phase for iteratively addressing and responding to user feedback across the suggested areas of digital behavioral health interventions (ie, appearance, behavioral prescriptions, burdens, content, delivery, message, participation, and assessment) [4].

The iSIPsmarter prototypes developed in this phase were paper- and web-based sketches that illustrated planned scenarios. The participants were sent links to the web-based prototypes, which included preprogrammed pages that displayed various feedback points as the participants moved through the content. This intentional rapid prototyping allowed the participants to interact with multiple prototype drafts and allowed the researchers to incorporate iterative feedback from the participants. The think-aloud method, which is a common approach to assessing the usability of digital health interventions and involves the participants verbally narrating their thoughts when completing a task related to the prototypes, was used. The researchers guided the process and asked the participants to complete specific tasks. Open-ended questions allowed for a robust understanding of usability and functionality. The interviewers documented where the participants encountered problems and difficulties using the prototypes.

This phase included 6 interviews to assess the participants’ comprehension and experience and the overall prototype functionality. The prototypes largely focused on the design and content of the iSIPsmarter user dashboard, daily SSB diary, action planning, and vignettes. The prototypes of the iSIPsmarter user dashboard and daily SSB diary were shown to the participants multiple times. The participants were asked to interact with the dashboard prototypes to collect data on their comprehension of the dashboard interface. This included navigating the dashboard to start a core (lesson) and add a daily SSB diary. Diary prototypes were also tested to collect data on the users’ understanding of adding daily SSB ounces. In addition, the participants were presented with SMS text message screenshots to assess their comprehension of daily SSB consumption tracking features. Similarly, prototypes of SSB and weight action plans were presented to the users to inform the development of a key intervention behavior change technique, personalized goal setting. Moreover, as previously mentioned, the journey maps were shared with the advisory team participants, and feedback was solicited to help narrow down, refine, and develop the vignettes associated with the personas.

End User Testing Phase

In this phase, the advanced functioning prototype or beta versions of the application are exposed to the end users and evaluated [16,18]. This phase can be carried out in a controlled laboratory setting, yet it is typically more useful when the intervention is field tested in the user’s own environment.

For iSIPsmarter, once the prototypes transitioned into the web-based RICE platform, the program-enabled website underwent numerous rounds of internal review by the CDT for clarity of content, flow, and transitions and for addressing
programmatic bugs. Subsequently, the end user testing phase began. The advisory team participants were given access to the iSIPsmarter website in their own environment. They completed the cores, including the embedded interactions, videos, and vignettes. Furthermore, they tracked their SSB intake and weight using the SMS text message feature and electronic cellular-enabled scale, respectively. A total of 3 semistructured interviews explored the usability and functionality of and user experience and satisfaction with the iSIPsmarter website.

Data Analysis

After each round of interviews, the researchers who conducted the interviews reviewed the audio transcripts and created interview summary documents that summarized each participant’s response to each question. Following each interview series, these summary documents with screenshots were shared with the CDT to incorporate feedback into the ongoing intervention content and programmatic development.

The analysis and data interpretation process involved several steps [61,62]. First, the 2 researchers who conducted the interviews reviewed the summary documents and independently identified key takeaway statements for each of the 13 interview rounds. Then, they met to build consensus and finalize the interview-level takeaway statements. Second, 2 researchers independently examined and deductively summarized the interview-level takeaway statements as higher-level findings and incorporated them into all 3 HCD phases, as aligned with the Model for Internet Interventions. During this second step, they also used CDT meeting minutes and artifacts from the development process (eg, prototype versions and drafts of core content) to inform higher-level findings. Again, the 2 researchers met to build consensus on the overarching phase-level findings. Finally, to illustrate the phase-level findings, key quotes from the advisory team members were extracted from the transcripts.

Results

Multimedia Appendix 1 illustrates the goals of the 13 semistructured advisory team interviews along with key interview-level takeaways. The phase-level findings, as aligned with the Model for Internet Interventions, are further summarized below.

Contextual Phase

Interviews 1, 3, 11, and 12 focused on the contextual and technology-related aspects of intervention delivery and data collection. The following key findings emerged:

- To promote participation, the participants described the need for built-in flexibility to account for varying levels of internet and SMS text message accessibility (digital health intervention use).
- The participants reported the need for accountability and personalized assessments.
- Plans for stepped care contacts were viewed as a helpful and important intervention component (support) that could boost adherence and core completion (behavioral prescription).
- The participants saw value in a weight monitoring component to promote reinforcement (participation) and personalization (assessment), yet barriers to weighing varied based on user characteristics and environmental factors, such as limited cellular service to transmit weights.
- The intervention enrollment procedures were easy to understand and complete.

These findings are illustrated by several key quotes from the advisory team members. One of the members expressed that they face difficulty in responding to SMS text messages owing to a lack of signal but that they would be able to email:

...I can receive her message but...I couldn’t answer her as prompt[ly] as I should or whatever. I’d have to find a place where I could get enough signal to send her back a reply. But with my home computer, if I get on there and have an email that I need to check, I can go ahead, you know, and go through and do it.

In reference to monitoring weight on the iSIPsmarter dashboard, one of the members said the following:

I think it’s a good reminder...to cut back on your sugary drinks...if you’re able to see your weight up there.

Prototype Testing Phase

The prototype testing phase allowed the CDT to iteratively address and respond to the user feedback. Interviews 2 and 8 focused on the dashboard and diaries; interviews 5, 6, and 7 were dedicated to the action planning concepts; and interviews 4, 5, and 6 focused on the personas. The final dashboard illustrated in Figure 2 contains a few examples of key intervention components, and a sample action plan is shown in Multimedia Appendix 2. For the persona-focused interview, 9 persona journey maps were shared with the advisory team participants, one of which is illustrated in Figure 3. On the basis of the participants’ insights into and reflections on the journey maps, the 9 initial personas were narrowed down to 6 and then further refined.

The main phase 2 findings are summarized as follows:

- The participants felt that flexibility in diary tracking methods and resources (eg, SMS text message, web-based tracking, drink cards, and paper diaries) would enable high engagement and participation among the users in tracking their SSB consumption and weight (behavioral prescriptions). For example, the participants who prefer paper diaries could first log their daily diary data via pencil and paper. Then, they could log into their dashboard weekly and back enter their diary data. Alternatively, the participants who prefer the digital methods could respond to the daily SMS text messages or email messages.
- Through multiple iterations, the dashboard became easy to navigate (low burden), with clear content, and contained helpful tailored user information (assessment), including visual cues to signify the completion of the core and diary tasks (appearance).
- The action planning process was easy to navigate (low burden); the content and delivery features were clear; a personalized and tailored feedback was perceived to be
helpful when setting and monitoring SSB and weight goals.

- Weight-related messages and content, including barriers and strategies, were relatable and easily understood.
- The vignettes were perceived as realistic and relatable, indicating an effective delivery approach for conveying key content.

Highlighting different patterns in SSB consumption behaviors among the vignettes was identified as a helpful messaging approach to improve the personalization and tailoring of content.

The following key advisory team member quotes support the findings:

Figure 2. Final dashboard screenshot.

In reference to the ability to navigate the action planning process, one of the members said the following:

*Everything else was pretty easy to understand. It was straightforward. It wasn’t lengthy as far as, like, a whole bunch of text that you needed to read. It provided really good examples for people to go by.*

In reference to the personas being realistic and relatable, the following was said:

*Kim because...She’s a single mom working full time. Busy schedule. Kind of just drink most of the day because she was too busy to eat...That was basically my story. So,…I can totally relate to that one. Wanted to be a healthy role model for her kids.*
End Users Testing Phase

Interviews 9, 10, and 13 focused on the end user testing. Table 1 highlights the final iSIPsmarter overview that was evaluated in this phase, including the asset summary, the user objectives, and a description of interactions or videos. The cores and interactions were developed and programmed in a manner that allows the users to go back and review the content and repeat the interactions to master skill development. Related to action planning, the underlying programmatic structure in cores 3 to 6 and the maintenance core are identical. However, the personalized feedback loops change based on the user’s diary data, progress toward their prior goals, and their new goals. The final SSB message bank includes 14 barriers and 80 strategies, and the weight-related message bank includes 13 barriers and 115 strategies. Therefore, each time a user completes an action plan, the personalized feedback has the potential to look very different (Multimedia Appendix 2).

On the basis of phase 2 findings, the CDT fully developed 6 vignettes and mapped key content through each of the iSIPsmarter cores and several interactions. Table 2 illustrates a final vignette, along with several examples of how the vignette is integrated across the cores.

The following phase 3 findings emerged when the participants were allowed to access the programmed intervention in their own environment:

- The overall appearance of iSIPsmarter was well received, including layout, organization, iconography, graphics, visuals, and color use.
- The participants reported high satisfaction with how the messages and content were delivered, including interactions, animations, videos, vignettes, and illustrations.
- Few burdens were reported, as the participants found the cores to be enjoyable, easy to navigate, user friendly, not overly text heavy, and of an acceptable length. Minor bugs and glitches were identified and resolved before launching the intervention.
- Participation was enhanced by the built-in flexibility for diary tracking.
- The behavioral prescriptions and personalization of assessments were well received, particularly the action planning process and automated emails and SMS text messages for reminding the participants to track their diaries.

A few key quotes from the advisory team member have been illustrated to support these findings. In reference to the interaction aimed at teaching and reinforcing the amount of sugar in drinks, one of the members said the following:

“When I clicked on...the container that you drink out of the most...and it showed you the little packets of sugar out there. I was like, Oh my gosh, when I seen that it really hit and I thought, No way do you need that.”

In reference to the personalized feedback, the following was expressed:

“I’ve always good to visualize because, you know, it’s hard to kind of think back to what I did you know last Thursday. But on something like this where I can see, OK, I had 12 ounces as last Thursday, I can kind of pull back in memory of the exactly OK, what I did those days that either lead me to only have this amount or led me to end up drinking this amount. So it...put those numbers into context a lot better than it would if you just said is average, this is your high this your low.”
In reference to the overall iSIPsmarter user experience, the following was said:

...[T]he website itself—that’s the thing that stood out the most to me. It’s well designed. It’s easy to navigate. It’s not too complicated. Everything’s well-organized...Everything is explained well, and it’s simple and concise and to the point. I like the level of the language that’s used. There’s not a whole lot of like jargon or extensive terminology or anything like that that might shy people away from participating. I think the readability of it is really well.

Table 2. An example vignette summary from iSIPsmarter.

<table>
<thead>
<tr>
<th>Name and image</th>
<th>Profile</th>
<th>Example stories throughout the cores</th>
</tr>
</thead>
</table>
| Beth           | • Beth is a 53-year-old married homemak-er and part-time church secretary. She has two grown kids who live away from home. She cares for her elderly father  
• Sugary drink pattern: Total ounces=64 ounces per day  
• 16 ounces frappe, three 16 ounces glasses of sweet tea throughout the day  
• Reasons for drinking sugary drinks: likes the taste; enjoys making sweet tea for family and friends; it’s a stress reliever  
• Physical activity: sedentary—has a gym membership but rarely goes; wants to be more active but has a hard time fitting it all in while caring for her father  
• Health status: just told she has diabetes and has gained 20 pounds in a year  
• Reason for joining iSIPsmarter: wants to manage the diabetes so she can be healthy, travel with her husband and take care of her family  
• Quote: “It’s just part of who I am”  
• Core 1—Texting Makes Tracking Easy: The daily text makes tracking my sugary drinks easy. I simply respond to the text with my total sugary drink ounces from yesterday. Also, I like seeing everything in one place when I sign in to iSIPsmarter.  
• Core 2—Getting My Diabetes Under Control: I was shocked when my doctor told me I had diabetes. My doctor told me getting to a healthier weight and making better food and drink choices were important to manage my diabetes. I started thinking about my health more. I know I have to take care of myself to keep taking care of my family. I saw that I drank too many sugary drinks and learned its impact on my dia-betes and weight. I knew cutting back on sugary drinks was going to be hard. Sweet tea and my frappes were just part of my routine.  
• Core 3—Getting Help from Others to Stay Accountable: I work part time for my church, and the staff often goes out to eat together. When eating out with my coworkers, it was harder to stick to my sugary drink goals. I told my co-worker Mary about iSIPsmarter and my goals and asked her to help me. Getting her support, helped me stay accountable.  
• Core 5—Could be Saving Money for College Tuition: My husband and I are helping our son with college costs so he doesn’t finish school with too many student loans. We are always looking for ways to save money. The frappes I was getting every day were $5 each day or $1,820 a year! This amount of money will definitely help cover some of his college costs.  
• Core 4—Getting Healthy Options from My Frappes: I know when I drink sugary drinks, it impacts my weight and my mood. I want to cut back on sugary drinks to eat healthier and be more active. I can get healthy options from my frappes that are healthier and less calorie intensive.  
• Core 5—Could be Saving Money for College Tuition: My husband and I are helping our son with college costs so he doesn’t finish school with too many student loans. We are always looking for ways to save money. The frappes I was getting every day were $5 each day or $1,820 a year! This amount of money will definitely help cover some of his college costs. |

Discussion

Principal Findings

This paper presents a flexible, step-by-step approach to and an HCD process for developing digital behavioral health interventions using iSIPsmarter as an illustrative example. By applying the Model for Internet Interventions [4] as well as best practices in HCD [16] and instructional design processes [17-20], we have been able to clarify assumptions and address key areas of the iSIPsmarter intervention that were hypothesized to support participants’ engagement and promote behavior change. As evidenced by the rich feedback received from the advisory team members, the human-centered methodology was instrumental in our development process. Likewise, the value of our robust process is exemplified by our resulting user-informed, high-quality products (eg, iSIPsmarter core components, vignettes, dashboard, and personalized action plan). Our approach can be interpreted within the context of the findings of recent narrative scoping review by Kip et al [14], in which a variety of methods (eg, interviews, focus groups, questionnaires, card sorting, and usability testing) and products (eg, prototypes, personas, and behavior change strategies) used for participatory eHealth development processes have been highlighted. Similar to the conclusions of Kip et al [14], our approach illustrates the importance of researchers and developers using the most appropriate methods to match their objectives and user characteristic context.

In terms of the iSIPsmarter development process, some of the biggest challenges faced when transitioning from the original group class structure to the digital structure include (1) replicating engagement provided by participant-to-participant and participant-to-instructor communication and relationships; (2) balancing the demands and cognitive load of the behavior change content (especially planning for the potential low health literacy skills of our targeted users) and trying to mimic visual, experiential, and hands-on class-based activities; (3) building flexibility to account for varying levels of internet and SMS text message accessibility among users; (4) creating a seamless user experience by integrating internet, SMS text messages, and sensor information; (5) operationalizing all aspects of the intervention content for digital delivery; and (6) automating all elements while ensuring that all permutations were considered. Similar challenges have been highlighted in a few other digital behavioral health intervention adaptation papers [14]. However, these issues are often overlooked and underreported in the literature, which may partly explain the low uptake of digital behavioral health interventions as well as the suboptimal fit between the characteristics of the technology and user needs. Simply converting evidence-based content traditionally delivered face-to-face to a web-based or digital format, without adequate attention to these and other challenges, is an insufficient approach to engaging the intended users or promoting improved health outcomes. Researchers and developers should rely on the established models and frameworks, such as the Model for Internet Interventions [4-9], to help anticipate and guide key
decisions when embarking on the development of digital behavioral health interventions.

Although several behavioral interventions targeting SSB consumption reduction among adults have been developed and evaluated [30], only a few have used scalable digital approaches. Moreover, only one other known adult-focused digital intervention targeting SSB consumption reduction has applied user-centered methodologies in the formative phases of intervention development [31]. Similar to our study, the study by Tonkin et al [31] focused on disadvantaged and nonurban adults, and its findings revealed the importance of understanding the available technology and patterns of its use as well as participants’ preference for stories, role modeling, and gamification, which foster engagement with the intervention. In addition, digital weight self-monitoring has become a cornerstone of many weight-related behavioral interventions, and greater adherence to self-monitoring is associated with better outcomes [63-65]. Unfortunately, consistency and disengagement in digital weight self-monitoring are known to be problematic [66]. Although several studies have investigated experiences of self-monitoring at the conclusion of interventions [67,68], there is a dearth of published studies that have applied HCD processes in formative intervention development stages to understand and build-in behavioral strategies to address potential personal and environmental barriers to digital self-monitoring [69].

On the basis of our experiences with SIPsmarter, we offer 6 broad considerations for other teams developing or adapting digital behavioral health interventions (Textbox 1).

In terms of study implications, the efficacy of SIPsmarter in reducing SSB consumption in rural Appalachian adults is currently being evaluated in a randomized controlled trial that includes a 2-group design (SIPsmarter vs static Participant Education website) with 4 assessment points (Clinical Trial Registry: NCT05030753) [36]. When efficacy and other summative data are available, they will provide additional insights to inform the potential value of applying an HCD process to build SIPsmarter as well as identify future areas of study.

**Textbox 1. Six recommendations for the development or adaptation of digital behavioral health interventions.**

1. **Assemble a multidisciplinary team of experts and end users:** similar to our SIPsmarter experience, the value of multidisciplinary team science and participatory processes in the development of digital behavioral health interventions is largely supported by other studies [15,16,18,35]. Our multidisciplinary content development team brought together expertise in the areas of nutrition content, behavior change, and rural health and worked alongside experts in digital behavioral health interventions, software engineering, instructional design, and user-interface design. Likewise, involving advisory team members with lived experiences in the targeted Appalachia region and with previous involvement in the SIPsmarter intervention brought immense value to the adaptation process. By applying human-centered design (HCD) principles [16], the advisory team members critically responded to iterative prototype versions, which helped shape key intervention decisions.

2. **Support efficient communication and decision-making processes among teams:** anticipating diverse feedback among different stakeholders, coordinating efficient communication among subteams, and finding a compromise are imperative to efficiently advance the HCD process [16,18]. For example, in our study, we coordinated communication and cooperation among 3 different subteams (ie, content, technology, and advisory teams) working to develop and advance SIPsmarter. In some instances, the advisory team requirements were different and contradictory to one another and to the requirements of the content development team.

3. **Define areas of adaptation at the onset of the process:** Similar to the adaptation of any behavioral intervention, the adaptation of digital behavioral health interventions can be driven by several distinct purposes. For example, the Adaptome [70] describes 5 potential sources of intervention adaptations: core components, culture, mode of delivery, target audience, and service settings. We adapted SIPsmarter from the evidence-based SIP smart ER trial with the clear goals of preserving the core components, cultural aspects, and the intended rural Appalachia target audience. This allowed us to concentrate on the mode of delivery and add key content around weight management. For example, we were able to focus end user feedback on potential digital content decisions in Appalachia [21-26]. This feedback guided us to build SIPsmarter with flexible features intended to enhance the likelihood of engaging in web-based cores, SMS text messages, and in-home weighing using cellular-enabled scales.

4. **Apply the available models and frameworks to guide digital intervention development processes:** Despite the promise and increased availability of digital behavioral health interventions, rapid disengagement and small effect sizes remain problematic [3]. Although evidence-based behavioral content is essential, it is only one of a multitude of factors that must be considered when developing or adapting digital behavioral health interventions. By applying the available models and frameworks to guide digital intervention development processes, researchers can identify and operationalize comprehensive components that affect the engagement and impact of digital behavioral health interventions [4-9].

5. **Clearly define instructional design goals to guide the HCD process:** To improve the likelihood of digital health applications meeting the needs and requirements of the users, instructional design objectives (eg, learning, affective, cognition, or psychomotor objectives) should be applied in the HCD process [15,16]. For example, in the SIPsmarter interviews, we were interested in evaluating the participants’ knowledge, problem-solving skills, attitudes, and values associated with completing certain tasks (eg, navigating the dashboard, completing an action plan, and engaging with an interaction). By having clear instructional design objectives, we were able to better understand and modify SIPsmarter features to meet the needs of the users and support the achievement of improved sugar-sweetened beverage consumption behaviors.

6. **Allocate appropriate resources and time to successfully execute HCD processes:** The time and resources required to develop or adapt digital health interventions can vary widely; however, they are often underestimated. For example, our iterative 3-phased HCD process was nearly an 18-month process. Researchers and developers should carefully consider and anticipate investment in robust HCD processes, as these processes are likely critical to the long-term uptake, engagement, and impacts of most digital health interventions.

**Limitations**

Overall, 3 key limitations of this study should be considered. First, generalizability may be limited by the relatively small sample of advisory team members, the targeted rural Appalachia region, and SSB-specific content. Second, qualitative interviews can result in socially desirable responses. However, efforts were
made by our trained interviewers to minimize this potential limitation by probing and clarifying the participant responses. This approach resulted in the participants providing thoughtful and constructive critiques of iSIPsmarter. Third, although we applied the Model for Internet Interventions, we limited the focus and scope of our manuscript to highlight the 8 main areas that comprise digital health applications [4]. Future efforts could focus more broadly on describing the other design-related components and elements that guided the iSIPsmarter adaptation process. Despite these potential limitations, we hope that the illustrated processes, scientific frameworks, and context-specific instructional design methodology will have wide applications in the development of digital behavioral health interventions across multiple geographic and behavioral contexts.

Conclusions
Our process emphasizes the value of researchers and developers applying the existing models and frameworks as well as best practices in HCD and instructional design processes in digital intervention development processes. Together, the complementary skills of the CDT and advisory team members were invaluable in the iSIPsmarter adaptation process. The importance of the contextual and iterative prototype testing phases was largely reinforced by the overwhelming positive feedback received in the user testing phase. By illustrating iSIPsmarter content, we have highlighted the user-informed, high-quality products that resulted from our robust HCD process.

Acknowledgments
The authors would like to acknowledge the time and dedication of our advisory team participants as well as members of our technology team, including Gabe Heath, Nicole Le, Ellia Hoepner, Ian Terrell, Alan Lattimore, Brendan Murphy, and Rom DuPlain. This study was funded by the National Institutes of Health, National Institute on Minority Health and Health Disparities (R01MD015033). The National Institutes of Health was not involved in the design of this study or writing of this manuscript.

Authors’ Contributions
JZ, LR, and KP were responsible for the acquisition of funding. JZ, LR, AR, CF, KP, and MH were engaged in all aspects of the 3-phased development process. AR and CF led the advisory team interviews and initial data coding. JZ and AR prepared the first draft of the manuscript. All the authors critically reviewed, edited, and approved the final version of the manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Goals and outcomes of the 13-series user-centered design process.
[DOCX File, 33 KB - humanfactors_v10i1e41262_app1.docx ]

Multimedia Appendix 2
Example of a user’s action plan.
[PNG File, 541 KB - humanfactors_v10i1e41262_app2.png ]

References


34. University of Virginia. URL: https://research.virginia.edu/ [accessed 2023-01-24]


https://humanfactors.jmir.org/2023/1/e41262 JMIR Hum Factors 2023 | vol. 10 | e41262 | p.63 (page number not for citation purposes)


Abbreviations

CDT: content development team
HCD: human-centered design
IVR: interactive voice response
RICE: Research Infrastructure Containing eHealth
SSB: sugar-sweetened beverage
Original Paper

Development of a Digital Assistant to Support Teleconsultations Between Remote Physicians and Frontline Health Workers in India: User-Centered Design Approach

Neha Verma¹, BE, MS, PhD; Harold Lehmann¹, MD, PhD; Amal Afroz Alam², BE, MSE; Youseph Yazdi¹, MBA, PhD; Soumyadipta Acharya¹, MBBS, MSE, PhD

¹Division of Health Sciences Informatics, Johns Hopkins University, Baltimore, MD, United States
²Intelehealth, Baltimore, MD, United States

Corresponding Author:
Neha Verma, BE, MS, PhD
Division of Health Sciences Informatics
Johns Hopkins University
3400 North Charles Street
Clark Hall, Suite 202
Baltimore, MD, 21218
United States
Phone: 1 4108707146
Email: neha@intelehealth.org

Abstract

Background: Many low- and middle-income countries have adopted telemedicine programs that connect frontline health workers (FHWs) such as nurses, midwives, or community health workers in rural and remote areas with physicians in urban areas to deliver care to patients. By leveraging technology to reduce temporal, financial, and geographical barriers, these health worker–to-physician telemedicine programs have the potential to increase health care quality, expand the specialties available to patients, and reduce the time and cost required to deliver care.

Objective: We aimed to identify, validate, and prioritize unmet needs in the health care space of health worker–to-physician telemedicine programs and develop and refine a solution that addresses those needs.

Methods: We collected information regarding user needs through ethnographic research, direct observation, and semistructured interviews with 37 stakeholders (n=5, 14% physicians; n=1, 3% public health program manager; n=12, 32% community health workers; and n=19, 51% patients) at 2 telemedicine clinics in rural West Bengal, India. We used the Spiral-Iterative Innovation Model to design and develop a prototype solution to meet these needs.

Results: We identified 74 unmet needs through our immersion in health worker–to-physician telemedicine programs. We identified a critical unmet need that achieving optimal teleconsultations in low- and middle-income countries often requires shifting tasks such as history taking and physical examination from high-skilled remote physicians to FHWs. To meet this need, we developed a prototype digital assistant that would allow FHWs to assume some of the tasks carried out by remote clinicians. The user needs of multiple stakeholder groups (patients, FHWs, physicians, and health organizations) were incorporated into the design and features of the task-shifting tool. The final prototype was shared with the health workers, physicians, and public health program managers who expressed that the tool would be useful and valuable.

Conclusions: The final prototype that was developed was released as an open-source digital public good and may improve the quality and efficiency of care delivery in health worker–to-physician telemedicine programs.

(JMIR Hum Factors 2023;10:e25361) doi:10.2196/25361

KEYWORDS
telemedicine; telehealth; eHealth; mobile health; mHealth; community health workers; frontline health workers; digital health assistant; task shifting

https://humanfactors.jmir.org/2023/1/e25361
Introduction

Background

Over 3.8 billion people, half the world’s population, lack access to essential medical care globally [1]. In low- and middle-income countries (LMICs) such as India, people living in rural areas must often travel far and spend a significant amount of time and money to access even basic medical care [2]. Despite the cost, they face long wait times and suboptimal quality of care at overburdened government facilities or have to navigate a fragmented private sector resulting in high out-of-pocket expenditures [2]. Several telemedicine initiatives have been implemented by different organizations in India [3-6] with unique considerations and implications for health care delivery. Prominent among these are the health worker-to-physician programs, which are defined as those where frontline health workers (FHWs) can facilitate teleconsultations for a patient with a licensed physician [7]. During teleconsultation, the health worker can take the history, examine the patient, and convey the findings to the physician [7]. They can explain and reiterate the physician’s advice to the patient [7]. The term FHW encompasses different cadres of health workers, including pharmacists, nurses, and volunteer community health workers [8], with each group having varying levels of training and experience [8]. The use of technology and mobile tools to support FHWs at the point of care is well studied in the literature for various use cases such as community-based information systems, electronic medical records (EMRs), learning and training systems, and telemedicine [8-10].

Through clinical immersion at 2 telemedicine clinics in 1 such project (Rural Health Kiosk, launched in 2015) in rural and remote areas in West Bengal, India, we studied the process of care delivery in health worker–to-physician telemedicine programs. The Rural Health Kiosk project aimed to reduce the challenges of geographical access to health care in the hinterlands of West Bengal. Two teleclinics were set up under this project in Barhata, a remote village, and Bali, an island in the Sundarban delta. These clinics are operated by local females from the village (called health assistants [HAs]) who have completed a government-accredited paramedical training program. HAs connect patients from the village to backend physicians who are based in Kolkata city using low bandwidth telemedicine technology for evaluation and decision-making in the clinical management pathway. They also conduct some diagnostic tests using low-cost point-of-care diagnostic sensors (eg, electrocardiogram and blood glucose). The physician communicates with the HA and the patient and accordingly prescribes medications and advice. This information is then transmitted to the HA who prints the prescription and explains and hands it over to the patient. Patients can purchase medicines via a pharmacy near the kiosk. In addition, the physician visits the kiosk once a month to administer care to patients that cannot be managed via telemedicine. Emergency patients are referred to the nearest secondary or tertiary care facility. Patients pay a small fee for the services, and additional revenue streams are generated through the sale of medicine, physiotherapy, and diagnostic pathology tests. The HA also provides community health care services such as immunization awareness, ante- and postnatal care, newborn care, and adolescent reproductive health awareness.

In a prior publication, we described the development of a software platform to support teleconsultations between FHWs and remote physicians [11]. In this paper, we describe the design and development of a key component of the software platform, a digital assistant to task shift history taking and physical examination to FHWs so that the remote physician has some initial information before beginning the teleconsultation. The use of digital assistants to improve the quality of clinical information gathered in a health care setting has been well studied in high-resource country settings [12-15]. We hypothesized that such a digital assistant could improve the quality of clinical information gathered in the teleconsultation process. It would have an added significance in an LMIC setting where physicians have to deal with high patient loads [16] and where diagnostic testing access is poor, leading to a greater reliance of physicians on the patient history to make a diagnosis [17].

Objectives

The objectives of this study were as follows: (1) to understand the needs of all stakeholders (patients, FHWs, physicians, and health organizations) during a telemedicine encounter between a rural patient and remote physician, which is facilitated by a health worker, and (2) to develop a solution to meet these needs. We used a bedside-to-bench-to bedside approach from the Spiral-Iterative Innovation Model for bioengineering innovation and design [18]. Many digital health projects do not move beyond the pilot stage, an important reason being that the needs and wants of stakeholders are not taken into account or that the system is not well designed for a local context [18-21]. Hence, we used a human-centered design approach and clinical immersion in the project location to incorporate the perspectives of stakeholders into the design of a solution and mitigate the risks of poor user adoption. We used the Spiral-Iterative Innovation Model [18] a biodesign approach to understanding, validating, and prioritizing unmet user needs in the health care space, as well as the creation, assessment, and refinement of solutions that address those needs (Multimedia Appendix 1 provides a detailed diagram of the model). The model requires designers to consider the perspectives and needs of all people and organizations impacted by the problem and its solution. These perspectives and their issues are organized into 4 domains—“clinical/medical, commercial/business, technical/design, and strategic/organizational”—that are evaluated during multiple iterations of the design process [22].

This paper describes the observations, insights, and unmet user needs that we identified. On the basis of these needs, we developed a prototype task-shifting digital assistant to support teleconsultations between FHWs and physicians, which was validated through user feedback. We believe that the findings of this study are highly relevant because access to qualified health care providers is a major challenge in rural areas, which has only been exacerbated by the COVID-19 pandemic. Travel restrictions, rising unemployment, and fear of visiting health facilities further prevent rural patients from seeking care when
and where they need it, lending greater importance to telemedicine.

**Methods**

We performed ethnographic research by direct observation and through key stakeholder interviews between August 2015 and June 2016 at 2 teleclinics of the “Rural Health Kiosk project” in West Bengal, India, implemented by JSV Innovations.

**Design Methods: The Spiral-Iterative Innovation Model**

We performed 3 design iterations using this model that are listed in subsequent sections.

**Opportunity Discovery**

The first step in the Spiral-Iterative Innovation Model process is the careful observation and extraction of actionable insights into potential stakeholders’ needs in the health worker–to-physician telemedicine encounter.

A team of 6 researchers (graduate students) performed the ethnographic research and shadowed health workers, physicians, and patients at various locations. We observed a total of 37 stakeholders (n=5, 14% physicians; n=1, 3% public health program manager; n=12, 32% community health workers; and n=19, 51% patients) during the opportunity discovery phase at various clinical sites. Clinical immersions were carried out at the following 5 locations: 2 teleclinics, 1 hub hospital, and 2 physician’s homes. Clinical sites were chosen to include all the possible locations involved in a health worker–to-physician telemedicine encounter (maximum variation sampling). We surveyed all health workers, physicians, and program managers involved with the Rural Health Kiosk project. The 12 health workers (6 at each teleclinic location) were females between the ages of 20 and 50 years. They lived in the communities served by these clinics. Their education level ranged from class 8 to graduate level. All health workers were certified “Home Health Aides,” a government-recognized allied health and paramedical training certificate program. Remote physicians were general physicians with an MD degree and between 20 and 40 years of clinical experience. The limited number of health workers and physicians would impact the generalizability of results; hence, we tried to limit our needs to those expressed during the health worker–patient-physician teleconsultation and not individually expressed opinions.

We used a purposive sampling approach for patients. The health workers and program manager invited community members to participate in an observed teleconsultation. The community members were invited such that their characteristics represented the types of patients who visit the teleclinic, including adult male and female patients (aged 18-65 years); children and adolescents (aged 0-18 years); older adults (aged ≥65 years); and patients who were living below the poverty line; had low literacy; belonged to scheduled castes, scheduled tribes, or other backward castes; and were farmers. We stopped recruiting more patients when saturation was reached, and very little new information was gained by the observation of a patient encounter.

We obtained oral consent from stakeholders before the interview or observation of the patient visit. The interviews were conducted in Bengali, Hindi, or English and were held at the clinical immersion site soon after the telemedicine encounter. They were individual interviews lasting between 30 and 60 minutes. The program manager (who was familiar with all 3 languages) served as a translator and was present for the observations and interviews. In addition, 2 members of the study team were fluent in Bengali and 2 in Hindi.

Observations were made through immersion into the clinical environment to study the behaviors, perspectives, and challenges of the stakeholders and factors in the user environment that influence and shape stakeholder behavior [18]. To ensure the completeness of capture of these observations, we sought multiple field settings, including rural telemedicine clinics, hub hospitals, and homes of remote physicians who would respond to cases while working from home. The observations were recorded with an observation code, date, location, persons involved, and a brief description of the activity being observed. After the teleconsultation, we conducted semi-structured interviews with the patient, health worker, physician, and program manager to understand the barriers and facilitators that they experienced in the teleconsultation process.

**Needs Selection**

The unmet needs identified through direct observation and interviews were analyzed using root cause analysis to arrive at the unmet user need. These were developed into unique “needs statements” using a standardized format consisting of a subject, a verb, a desired outcome, and optionally additional context. For example, “A doctor (subject) needs a way to remotely collect (verb) a patient’s clinical information (desired outcome) in order to provide an accurate diagnosis and treatment plan (context)” [22,23]. Each needs statement was provided a code number and linked with the observations from which it was derived. Issues from each of the 4 quadrants were considered from each stakeholder’s perspective. Using the 4 domains, we clustered user needs into thematic areas, filtered them, and prioritized them to select the most critical need to be addressed.

**Solution Design**

We developed a prototype of the solution to meet the need informed by all 4 perspectives. The prototype solution was shared with the users who were observed to understand if the design was acceptable and met their needs. This was done through user feedback interviews with the community health workers (2 group settings with 6 community health workers in each group) and individual interviews with each of the physicians and the public health program manager. The feedback from the participants was analyzed and incorporated into the final design of the solution.

**Ethical Considerations**

This study was approved by the Johns Hopkins Institutional Review Board and the JSV Innovations Institutional Review Board in India, and informed consent was obtained as per protocol (IRB00050927).
Results and Discussion

In this paper, we present a combined Results and Discussion section because of the iterative nature of the design method used. The results and the corresponding discussion of the significance of these results for each design iteration are presented in subsequent sections.

Iteration 1: Opportunity Discovery

We identified 74 unmet needs at various stages of the teleconsultation workflow, which were developed into needs statements. These needs have been clustered into 11 thematic areas and presented along with example needs statements in Table 1. Multimedia Appendix 2 provides the complete list of needs statements.

Table 1. Selected needs statements identified during opportunity discovery, grouped into thematic areas.

<table>
<thead>
<tr>
<th>Code</th>
<th>Thematic area(^a)</th>
<th>Example needs statements(^b)</th>
</tr>
</thead>
</table>
| MI   | Medical information (n=14) | • “Doctors need a way to get accurate medical information (signs and symptoms) about the patient in order to provide correct diagnosis and treatment plan” [MI.01]
|      |                      | • “FHWs need a way to accurately record patients’ medical information (signs and symptoms) and share it with the remote doctor to improve the quality of diagnosis and treatment.” [MI.02] |
| KC   | Frontline health worker knowledge and competency (n=14) | • “Health organizations need to standardize the skills of FHWs so that all kiosks can provide quality care.” [KC.12] |
| ID   | Instrumentation and diagnostics (n=9) | • “Doctors need high quality, accurate stethoscope results (heart and lung sounds) to effectively diagnose patients.” [ID.04] |
| MC   | Medications and compliance (n=4) | • “Doctors and FHWs need a way to improve patient compliance with medical advice and medications to improve patient outcomes.” [MC.01] |
| IC   | Patient education and informed consent (n=4) | • “FHWs need a way to explain concepts of informed consent (such as risks and benefits of telemedicine and data privacy) in a manner that is comprehensible to rural patients with low literacy backgrounds so that patients can make informed decisions.” [IC.02] |
| PX   | Patient experience (n=4) | • “Patients need to feel like their information is accurately conveyed to the Doctors to engender trust in the kiosk model and increase patient acceptance.” [PX.04] |
| EC   | Emergency care (n=4) | • “Health organization, FHWs, and doctors need to better identify patients needing emergency services to provide first aid, stabilize the patient, and promptly initiate a referral.” [EC.01] |
| CW   | Clinical workflows (n=6) | • “Health organizations need to optimize patient flow, reduce process redundancies, and increase patient throughput to improve teleconsultation efficiency.” [CW.01] |
| CO   | Communication (n=5) | • “Patients, FHWs, and doctors need to communicate in a language that is comfortable for all stakeholders in the telemedicine interaction.” [CO.01] |
| TE   | Telecommunications infrastructure (n=3) | • “Health organizations, FHWs, and doctors need technology to function reliably, including during power outages and periods of low/no internet.” [TE.03] |
| EU   | Ease of use (n=2) | • “Doctors and FHWs need technology which does not consume a lot of time in data entry and fits seamlessly into the clinical workflows.” [EU.02] |
| FS   | Financial sustainability (n=5) | • “Health organizations need to have a very high operational efficiency to achieve sustainability and scalability of the model.” [FS.01] |

\(^a\)The number in parentheses includes the total number of needs statements in that thematic area. For a complete list, see Multimedia Appendix 2.

\(^b\)For the complete list of example needs statements (a total of 74), see Multimedia Appendix 2.

Iteration 2: Needs Selection

Task Shifting in Telemedicine

Through interviews with the remote physicians, we recognized that there was a limitation on the types of patients for which a management plan could be remotely developed, owing mainly to the lack of trustworthy elicitation of the patients’ signs and symptoms. The accurate collection of medical information and its communication between the various stakeholders is essential to arrive at the correct diagnosis and treatment plan for the patient. Overall, the more relevant data the FHW can reliably collect from the patient via history taking, physical examinations, and point-of-care diagnostic sensors, the higher the ability of the remote physician to diagnose accurately. After evaluating each need on the 4 quadrants using the evaluation parameters (Multimedia Appendix 1), we identified the need...
for task shifting clinical information gathering to an FHW as a top need (MI.01; Table 1).

Significance of Task Shifting Information Gathering to an FHW in a Telemedicine Setting

In telemedicine, the remote physician is not in the same location and cannot directly see, touch, or hear the patient. The FHW serves as the “ears, eyes, and hands” of the physician [11]. The FHW does not have the required training or skill to be able to collect this information accurately. Collecting a comprehensive medical history and performing a clinical examination, the main pillars of arriving at a provisional diagnosis, require medical knowledge, training, and experience. The FHW typically cannot be trained comprehensively in these skills (short of going through medical school and a residency program). Hence, most telemedicine programs involve the FHW simply registering the patient with basic demographic details, serving as a telecommunications operator, and the remote physician does the remaining patient interview processes over the phone or a video call.

Task shifting history taking to other types of health workers or directly to the patient for self-reporting with the use of a digital assistant or a computer-assisted history-taking system to improve the quality of clinical information gathering has been well established in the literature, mainly being used in health care settings in high-income, developed countries [12,13]. A review of the literature reveals some clear benefits such as the improvement in documentation, reduction in time spent by a provider in documentation, the ability to collect more comprehensive and relevant information, and improvement in the quality of information gathered [12]. The drawbacks include the inability to capture nonverbal communication, frustration felt by users if the questions do not fit the scenario, user interface challenges, and irrelevant questioning [12].

It is essential to understand the role of patient history, physical examinations, and investigations in arriving at a diagnosis. A study in India with 100 in-person outpatient consultations showed that in 78.58% of the cases, the patient history led to the diagnosis [17]. In 8.17% of patients, the physical examinations led to a diagnosis, and in 13.27% of patients, investigations led to a diagnosis [17]. The study also showed that the physician’s confidence in the correct diagnosis increased subsequently at each stage of the clinical information–gathering process from 6.36 after history taking to 7.57 after physical examinations and to 9.87 after investigations as measured on a Likert scale from 1-10 [17].

A retrospective analysis of 32 malpractice suits in telephone consultation–related adverse events showed that poor documentation (88% of the 32 cases) and faulty triage decisions because of incomplete history taking over the phone (84% of the 32 cases) were the leading causes of diagnostic error [24]. A study by Resneck et al [25] observed history taking to be rushed or incomplete in direct-to-patient web-based consultations because of time pressures as well as the physician not being able to see the patient. Simple relevant questions related to history taking, including allergies, medications, or pregnancy status, were routinely missed, resulting in missed diagnoses [25].

In rural India, access to diagnostic laboratories is poor, necessitating additional patient travel to conduct basic laboratory tests. We observed that the physician often has to rely on a well-taken history to arrive at a diagnosis and management plan. This places added value on the patient interview because it is often the primary basis for decision-making.

Although the patient interview is an essential component of the diagnostic decision-making process, history taking in resource-limited settings may often be rushed or incomplete because physicians are overburdened. A systematic review of the average time taken for a primary care consult in 71 countries showed a wide variation in the average time taken for a consult between developed and developing countries [16]. For example, an average primary care consult in India lasts 2.5 minutes and an average primary care consult in Bangladesh lasts 48 seconds [16].

A study in rural India used standardized patient actors to assess the quality of care of health care delivery by 224 public sector and private sector providers (qualified and unqualified). It analyzed the care quality in terms of consultation length. In addition, the study used a checklist of essential history taking and examination steps that providers should follow during a consultation and evaluated what percentage of these steps were actually followed by the providers. The average public sector provider-patient interaction lasted 2.4 minutes during which the provider completed 16% of the checklist items, and the average private sector provider-patient interaction lasted 3.7 minutes and gathered 22% of information from the essential checklist [26]. The consultation length was strongly correlated with the completion of more items on the checklist [26]. We observed similar concerns through our ethnographic research. The time taken by the physician to complete a teleconsultation is often a function of the patient load, which is very high in resource-limited settings. Optimizing the amount of time spent by a physician on a teleconsultation, without compromising the quality of care, has important implications for the financial sustainability and scalability of telemedicine in LMICs.

Iteration 3: Solution Design and Development

Overview

We integrated perspectives from all 4 quadrants when designing the prototype. We developed multiple versions of the prototype and refined them. Insights from other needs statements were also incorporated into the design so that the overall solution could address several unmet needs. These are summarized in Figure 1.
The final prototype was a “digital assistant” that enables the task shifting of history taking and physical examination tasks to FHWs to allow for effective teleconsultations. The goal of this assistant was not to provide a final diagnosis but to guide the FHWs to collect comprehensive patient information to share with the remote physician.

**Workflow**

The FHW uses a mobile app with the digital assistant for elucidating clinical information from patients and also conducts some diagnostic tests using point-of-care diagnostic devices (blood pressure, blood sugar, electrocardiogram, etc.). The combined data are sent to a remote physician using low bandwidth internet for evaluation and decision on the clinical management pathway. The physician, on reviewing the case, communicates with the FHW and the patient over a phone call or a video call for further clarification or examination and accordingly prescribes medications and provides advice. This information is then transmitted to the FHW who prints the prescription, explains, and hands it over to the patient. Patients can purchase some basic medicines from a pharmacy nearby. Emergency patients or patients who cannot be managed via telemedicine are referred to the nearest secondary or tertiary care facility. The workflow of a teleconsultation when guided by the digital assistant is shown in Multimedia Appendix 3.

**Mobile App**

We developed the task-shifting tool as a mobile app because of the growing use of low-cost mobile devices in global health care delivery. Medical history taking and physical examination are core clinical skills taught to physicians and often require years to attain proficiency. As FHWs do not have the medical knowledge and training to take a complete, evidence-based medical history or to conduct examinations as a qualified physician would, a mix of job aids was built into the tool to guide them contextually in what to ask and what to examine. We also developed a training protocol for the use of the tool. We developed the interface keeping in mind the need to promote improved confidence and increased capacity in FHWs to execute these skills at an adequate level of proficiency.

A sense of trust between the physician and the health worker is essential to their functioning effectively together as a care team. The fact that an evidence-based, knowledge-enabled digital tool to facilitate task shifting in a high-quality, standardized manner is behind the FHWs’ workflow can increase this trust while enhancing the quality of the information. The output note shared with the physician was concise and easy to read to enable efficient communication. It aimed to have all the information that is necessary and sufficient to arrive at a differential diagnosis. Irrelevant information that serves as “noise” for the physician, making it harder to focus on the information relevant to the case, was minimized.

The primary language of communication among the patient, FHW, and physician is the local language. This placed additional considerations on the use of terminology and proper translations of medical terms into the local language (Bengali). Medical terms were presented in a simple language that was easy for both the FHW and the patient to comprehend. History-taking questions were contextualized to local norms. We hypothesized that more time would be spent on history taking and examination when conducted by an FHW with the digital assistant instead of by a physician.
of the 2 to 3 minutes spent by a busy physician in an overcrowded outpatient clinic. This would have a profound impact on improving the patient experience. Overall, the solution was designed keeping in mind the need to meet the sociological and psychological goals of the patient interview, that is, responding to patients’ emotions and influencing their behavior. Data collected by the mobile app are stored in an open-source EMR system (OpenMRS) to ensure that each patient’s clinical information is tracked longitudinally and tied to a unique health identifier.

Protocol Development: Defining What Should Be Task Shifted and What Can Be Task Shifted

While building the protocols, it was pertinent to determine what could be safely and effectively task shifted to FHWs:

- Should abdominal palpation to elicit superficial tenderness be task shifted to an FHW?
- Should the collection of sexual history be task shifted to an FHW?
- Should the measurement of blood pressure be task shifted to an FHW?

The decision about what to task shift depends on the current education and skill level of the FHW; the amount of time and complexity required to train them in new skills; their ability to retain those new skills; and the feasibility of executing continuous training, learning, and competency assessment. The return on investment on training the FHW versus the value of the specific information to make a diagnosis given the context drives the decision of whether a specific skill should be task shifted. The patient’s willingness to share this information with the FHW when asked is an important consideration for deciding whether it can be task shifted. Thus, the need to collect specific data should be balanced with the acceptability and trainability of the health worker to gather it accurately.

Clinical Value of the Symptom or Sign in Making a Diagnosis

Key symptoms and signs allow physicians to rule in or rule out a diagnosis. The value of a symptom or sign is usually thought of in the context of making a diagnosis. The patient interview progresses in such a way as to arrive at a differential diagnosis. After this, the physician can order further tests to confirm a diagnosis or pursue a therapeutic pathway by making the best decision from the available data. Red flag symptoms that rule in diagnoses that would result in death or severe disability are “high value” because they allow the remote physician to identify patients needing urgent care or referral. Thus, the clinical value of a symptom or sign is intrinsically related to the morbidity of the diagnosis it can point to, its sensitivity, specificity, negative and positive predictive value, and its contribution toward deciding the appropriate therapeutic plan (Figure 2).

Figure 2. Framework for prioritizing what signs and symptoms are included in the task-shifting digital assistant with examples. We selected clinical information that is of high value and low effort to train a frontline health worker (FHW), with high patient and FHW acceptability as suitable targets for task shifting. Both the trainability and the clinical value are dependent on the baseline skill level of the FHW (nurse vs midwife vs community health worker) and the socioeconomic context of the clinic, such as its distance from the nearest diagnostic center. BP: blood pressure.
**Level of Effort Needed to Train the FHW in Capturing the Symptom or Sign**

Task shifting of data gathering for each sign or symptom is associated with a level of training that needs to be provided so that the FHW has the skills required to collect it accurately. For example, asking a patient since when they have had back pain and then selecting the duration of the pain does not require a lot of training, especially if the tool prompts the question and the FHW does not have to remember the context in which the question needs to be asked. Measuring vital signs using point-of-care devices requires a higher level of training. Examinations such as chest auscultation or palpating the margins of the liver require an extremely high burden of training and practical experience (Figure 2).

**Acceptability to the Patient**

From direct observations of multiple client-provider health interactions, we noticed that collecting information about sexual history for men by female health workers or questions about mental health issues may not be considered culturally appropriate for an FHW to probe as they are members of the same community. Furthermore, we hypothesized that acceptability relates to several aspects of the FHW—their gender and social status relative to the patient, the perception of their expertise in the eyes of the patient, and the prevalent sociocultural norms.

**Acceptability to the Health Worker**

We also observed that task shifting creates an additional workload for the health workers. Furthermore, FHWs in LMICs are often not well-compensated and are overburdened, leaving little incentive to take on other tasks. We observed that the health worker’s confidence in being able to use a digital tool or devices can also limit acceptability.

**Regulatory Considerations**

From the health organization’s perspective, such a tool would allow for standardization in task shifting so that FHWs with varying skills and abilities could perform consistently at high quality. An important consideration for health organizations is in assuming the regulatory uncertainty around task shifting to FHWs because the guidelines are often not laid out. An evidence-based approach to the development of protocols was adopted to minimize patient safety risks and minimize regulatory risks. Furthermore, although data and interoperability standards are currently in a nascent stage in India, as government regulations and standards for telehealth systems are adopted, the tool would also need to adhere to these standards to achieve integration with other similar vertical systems and prevent data from being siloed. Hence, we chose to use OpenMRS as an EMR backend for the system so that interoperability requirements could be achieved via the OpenMRS Fast Healthcare Interoperability Resources, Health Level 7 APIs, and the use of data dictionaries such as Systematized Nomenclature of Medicine Clinical Terms, Logical Observation Identifiers Names and Codes, and International Classification of Diseases, ninth and tenth revisions, already built into the OpenMRS platform architecture [27].

We built a committee of 11 members across all the stakeholder groups (distinct from the stakeholders observed in phase 1) to develop and review the protocols. Task-shifting protocols were developed from known evidence bases such as medical textbooks or guidelines issued by the health ministry with clear citations to the source of the protocol. Different members of the committee participated in the development as it related to their background and expertise (Table 2).
Table 2. Process of knowledge acquisition to develop task-shifting protocols to collect patient informationa.

<table>
<thead>
<tr>
<th>Step</th>
<th>Stage</th>
<th>Result</th>
<th>Committee members involved</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1: physician</td>
</tr>
<tr>
<td>1</td>
<td>Identify symptom list to cover the scope of most prevalent presenting complaints through literature review</td>
<td>67 presenting complaints identified</td>
<td>✓</td>
</tr>
<tr>
<td>2</td>
<td>Create data collection questionnaires to collect subjective data for the presenting complaints through a literature review and synthesis of evidence-based guidelines</td>
<td>67 data collection questionnaires compiled</td>
<td>✓</td>
</tr>
<tr>
<td>3</td>
<td>Identify simple physical examinations to collect objective data and map them to complaints</td>
<td>143 examinations identified</td>
<td>✓</td>
</tr>
<tr>
<td>4</td>
<td>Contextualization of questionnaires to the etiology and epidemiology of disease in India</td>
<td>67 questionnaires contextualized</td>
<td>✓</td>
</tr>
<tr>
<td>5</td>
<td>Feasibility assessment to remove history-taking questions and physical examinations that are difficult to task shift to health workers or have a high burden of training</td>
<td>Questionnaire list reduced to 51; examination list reduced to 93</td>
<td>✓</td>
</tr>
<tr>
<td>6</td>
<td>Translation of content into local language (Bengali) and adaptation to improve comprehensibility for patients</td>
<td>Translations complete and verified; 51 questionnaires and 93 physical examinations modified</td>
<td>✓</td>
</tr>
<tr>
<td>7</td>
<td>Adaptations to local social and cultural contexts</td>
<td>Adaptations complete and verified; 51 questionnaires and 93 physical examinations modified</td>
<td>✓</td>
</tr>
</tbody>
</table>

Physicians were responsible for the curation of medical knowledge to build the protocols. Public health program managers were experts in working with community health workers as well as in representing patient needs. The community health workers represented the sociocultural norms of the community that they served and the local dialect as well as patient needs.

User Feedback Interviews

The first version of the prototype was shared with the health workers, physicians, and public health program managers who were observed in the ethnographic design stage for feedback. The users saw a demo of the tool and directly interacted with the app for 1 hour. All the participants agreed that the digital assistant would be a useful addition to the telemedicine program and improve key project bottlenecks. The physicians and the public health program managers felt that task shifting the patient interview to an FHW may save the physician’s time and increase their ability to diagnose remotely. It could lead to better documentation of consults and increase the availability of actionable data for public health analysis. A physician observed that structured information in the output note can be used to trigger physician job aids such as differential diagnosis checklists and standard treatment guidelines.

Community health workers felt that the tool could improve their ability to interact with the patient and reduce back and forth communication with physicians. It would also increase the patient’s trust in their skills and in the program overall. They were concerned about whether they would be able to use the tool correctly and provided feedback to improve the user interface and make it simpler. All users expressed that the community health workers would need to be trained properly to use the tool effectively. Some users expressed concern that this may make the patient encounter too lengthy because of the detailed nature of the questioning workflows. The public health program manager observed that such a tool also can potentially implement further evidence-based medicine approaches to...
improve clinical care delivery in health worker–to-physician telemedicine programs. Owing to the small sample size, these findings cannot be generalized.

The first prototype thus addressed several needs identified in the opportunity discovery phase and gave us insight into concerns. Accordingly, the final version of the app that was developed incorporating the user feedback has been shown in Figure 3.

We developed the prototype version with history-taking questionnaires for 51 presenting complaints and 93 physical examinations (Figures 3A and 3B). We released the app source code under the free and opensource Mozilla Public License 2.0 [28]. Supporting documentation is provided on the internet [29].

Figure 3. (A) The user interface of the prototype digital assistant. The frontline health worker (FHW) can select the presenting complaints, answer detailed questions for each complaint, and collect past medical history and family history using a standard protocol–based adaptive questionnaire. The FHW collects vital signs and is guided through conducting physical examinations. (B) An example output note generated by the FHW using the digital assistant. The assistant also guides the FHW to capture images to share with the physician so that the physician may arrive at a diagnosis.

Limitations and Strengths
Many of these conclusions are based on a design ethnography process. A different set of user needs may have been discovered by a different set of researchers. The results are derived from observations in 2 communities in a single state in India. One has to be cautious in interpreting the generalizability of this approach to multiple geographies. Although the participants
reported both positive and negative feedback about the usefulness of the digital assistant, there is a high chance of positive bias in the user feedback interviews. The responses cannot be generalized because of the small sample size. Further validation of these hypotheses is required to see if these needs and the resulting solution can be applied broadly.

Although many digital assistants for history taking have been developed in high-resource settings, to the best of our knowledge, no such system has been designed comprehensively with the needs and requirements of a rural health worker-to-physician community-based telemedicine program in a resource-constrained environment.

Conclusions
A digital tool for task shifting clinical information gathering to an FHW has high significance and value in a telemedicine setting in an LMIC as observed through interviews with key stakeholders. We identified the key value propositions and user needs and presented a prototype of a task-shifting tool for telemedicine settings in LMICs. In a developing country setting, such a tool’s significance may be much higher, given the resource constraints that physicians operate under. The final prototype incorporated unique value propositions for all stakeholders—physicians, FHWs, patients, and public health program managers—and could result in an overall improvement in the quality of care delivered via telemedicine in resource-constrained environments. The prototype version was acceptable to the users. Future scope for development of this tool would involve additional iterations of the spiral innovation approach with the refinement of the prototype, testing, regulatory compliance, pilot implementation, field evaluation, and commercial validation. Additional research needs to be conducted to evaluate the digital assistant and its impact on various aspects of a telemedicine program such as its feasibility of implementation, impact on the diagnostic outcome, impact on improving health worker capacity and competence, patient and provider satisfaction, clinical safety, program quality, and efficiency of care delivery.

Acknowledgments
The authors would like to acknowledge JSV Innovations, Kolkata, India, for providing access to clinical field sites. The authors would also like to acknowledge Emily Eggert, Sean Mattson, and Elizabeth Lebling for their contributions to the ethnographic research.

Conflicts of Interest
NV is a founder of and serves as the chief executive officer of Intelehealth, a 501(c)(3) nonprofit supporting the development and implementation of the telemedicine software. SA is a founder of and serves as the Board President of Intelehealth. NV and SA are also the inventors of the technology involved in the Intelehealth app, which was used in this study. This arrangement has been reviewed and approved by the Johns Hopkins University in accordance with its conflict of interest policies.

Multimedia Appendix 1
The Spiral-Iterative Innovation Model that guided the 3 phases of (A) opportunity discovery, (B) needs selection, and (C) solution design.

Multimedia Appendix 2
Needs statements identified during opportunity discovery, grouped into thematic areas.

Multimedia Appendix 3
Swimlane diagram with the workflow of a patient–health worker–physician teleconsultation guided by a task-shifting tool.

References


Abbreviations

**EMR**: electronic medical record

**FHW**: frontline health worker

**HA**: health assistant

**LMIC**: low- and middle-income country
Evaluating Clinician Expectations of mHealth Solutions to Increase Rapid-Screening for HIV and Hepatitis in Migrant Populations in France: Qualitative Study

Carter Brown¹,², MPH; Guillaume Roucoux¹,³, MSc; Svetlane Dimi⁴, MD; Saleh Fahmi¹, MSc, MD; Raj-Banou Jeevan¹, PsyM; Olivier Chassany¹,³, MD, PhD; John Chaplin⁵, PhD, C Psych; Martin Duracinsky¹,³,⁶, MD, PhD

¹Patient-Reported Outcomes Unit, UMR 1123, Inserm, Université de Paris, Paris, France
²Pierre Louis Doctoral School of Public Health, Sorbonne University, Paris, France
³Unité de Recherche Clinique en Économie de la Santé, Assistance Publique – Hôpitaux de Paris, Hôpital Hôtel-Dieu, Paris, France
⁴International Vaccination and Travel Medicine Center, Centre de Santé Familia Sol, Creil, France
⁵Institute of Health and Care Sciences, Sahlgrenska Academy at Gothenburg University, The University of Gothenburg Centre for Person-Centred Care, Gothenburg, Sweden
⁶Service de Médecine Interne et d'Immunologie Clinique, Hôpital Bicêtre, Le Kremlin-Bicêtre, France

Corresponding Author:
Carter Brown, MPH
Patient-Reported Outcomes Unit, UMR 1123
Inserm
Université de Paris
1 Parvis Notre Dame - Place Jean-Paul II
Paris, 75004
France
Phone: 33 0140274664
Email: carter.brown@inserm.fr

Abstract

Background: Migrants underuse screening opportunities for HIV, hepatitis B, and hepatitis C despite elevated risk factors for contracting these infections. Language barriers are an often given as reasons for limiting access to services. Translation and communication apps increase communication and overall patient satisfaction in the patient-provider relationship. In the development and adoption of new technology, expectations play an important role.

Objective: This study aimed to explore health care professionals’ opinions and attitudes regarding their screening practices with migrants and their expectations for a new communication tool that could improve migrants’ screening use.

Methods: In this qualitative study, a purposive (diverse) sampling method was used to invite doctors and nurses who conduct rapid screening tests with migrants from 4 centers of the French Office of Immigration and Integration in 3 geographic regions of France. Semistructured interviews were conducted to survey their opinions on the rapid testing of migrants, the use of telephone interpreters, the concept of health literacy, and their expectations of a new communication tool that could overcome language barriers and promote rapid screening in the new migrant population.

Results: In all, 20 interviews were conducted with 11 doctors and 9 nurses with a median age of 58 (range 25-67) years. Participants favored the integration of an innovative communication tool in the context of rapid screening of migrants. However, there were concerns related to the implementation and added value of the tool while migrants were already reluctant to be screened. Expectations were for a tool that would present information in simplified French or a chosen language but also supports a positive attitude toward screening. Health professionals also expressed the wish that the technology could help with the collection of health data.

Conclusions: Feedback from health professionals provides a better understanding of potential formats, characteristics, functions, content, and use of an innovative, digital method to communicate with migrants with limited French proficiency. Findings contribute to the conceptual development of an electronic app and its implementation within the ApiDé study, which aims to validate a digital app to address language barriers to increase the use of screening among migrants with limited French proficiency in France.
Introduction

In 2018, nearly 6200 people were diagnosed with HIV in France. Among these individuals, more than half (3224/6200, 52%) had never been tested for HIV in their lifetime, 81% (5022/6200) were injection drug users, and 65% (4030/6200) were from sub-Saharan Africa [1]. A French survey from blood donors estimated the prevalence of hepatitis B virus (HBV) to be 53.1% (250/471) among migrants from endemic regions while attributing nosocomial exposure as the leading factor for hepatitis C virus (HCV) [2]. Furthermore, many HIV-positive migrants in Europe acquire their infections after migration [3].

The French health authority (Haute Autorité de Santé) guidelines for annual screening recommends that people who have multiple partners from endemic regions, notably sub-Saharan Africa and the Caribbean, should be screened regularly for HIV, HBV, and HCV [4]. Furthermore, recommendations suggest that these 3 tests should be conducted at the same time [5]. Between 2017-2020, a total of 21,133 migrants were tested for HIV, HBV, and HCV at the French Office of Immigration and Integration (Office Français de l'Immigration et de l'Intégration [OFII]) [6].

In France, legal migrants must undergo a medical examination as part of the administrative process to obtain a residence permit at OFII. During this consultation, health professionals (HPs) propose a free-of-charge, rapid antigenic and/or antibody diagnostic test (test rapide d'orientation diagnostique [TROD]) for HIV, HBV, and HCV with results in 20 minutes or less. Individuals who test positive are oriented for follow-up care, which is also provided free of charge.

Barriers to accessing screening services in the migrant population in France have been previously documented. HPs frequently cite linguistic barriers as major obstacles negatively impacting the acceptability of the diagnostic testing [7]. A recent study in immigrants in Canada found that language barriers interfered with preventative and screening services and ultimately lead to poor health outcomes [8]. Additionally, a study investigating knowledge, behavior, and practices related to HIV and sexually transmitted infections among migrants from sub-Saharan Africa living in Germany found that German language proficiency was one factor associated with knowledge about German HIV policies and HIV testing [9].

A systematic review investigating language barriers in migrant health care found that translation apps enable better communication in the patient-provider relationship and reduce overall consultation times [10]. Another study demonstrated that a mobile translation app contributed to the use of interpretation services and resulted in a high level of satisfaction among HPs [11]. The benefits that mobile health apps could provide to HPs' communication needs have not been investigated in relation to virus testing in migrants who have limited French proficiency (LFP).

Innovative methods to reach vulnerable and migrant populations are needed to increase the acceptability of TROD [12], but there is a lack of evidence on how HPs experience language barriers in this context. Knowledge of HPs’ expectations is fundamental to the development of new technology and plays an important role in determining the rate at which it will be adopted [13]. Therefore, it is important to explore expectations in the early phases of the new technology’s life cycle when there is uncertainty regarding performance [14].

This research is part of the STRADA study that started in 2017 to determine the acceptability of TROD from both the migrants’ and HPs’ perspectives [7]. The objective of this qualitative study was to explore how HPs envisioned a hypothetical new tool that could help increase communication with migrants with LFP to explain the importance of rapid screening at OFII. We believe that by engaging HPs in the conception of a future app that the app will be better made and that HPs would be more likely to use said app.

Methods

Population and Setting

This qualitative study was conducted in a population of doctors and nurses who work in 4 OFII centers in France (Lyon, Nice, Cergy, and Montrouge). Each center employed more than 5 HPs on site and had varying volumes and origins of migrants. A purposive (diverse) sampling method was used to include equal numbers of doctors and nurses, male and female, who regularly offer rapid screening test to migrants and who previously participated in the STRADA screening study. All interviews were conducted face to face in the informant’s workplace.

Interviews

Semistructured interviews were conducted using an interview guide (Multimedia Appendices 1 and 2). This guide was created with themes that emerged from previous interviews with migrants, data that have not yet been published, and with reference to literature and expert opinions. The interviewers (SF and RBJ) explored the HPs’ experience with migrants outside of the OFII context, medical visits with migrants at OFII, rapid testing of migrants, the use of telephone interpreters, the concept of health literacy, and the HPs’ opinion on the creation of an electronic tool to promote rapid screening of migrants with LFP. This paper solely covers the last theme; future articles will appear on the other topics.

Data Collection

In all, 20 in-person interviews were conducted from May 15 to October 20, 2019. The interviews lasted roughly 30 (range 15-44) minutes. Interviews were audio recorded. Data were
collected according to Consolidated criteria for reporting qualitative research (COREQ) guidelines [15] (see Multimedia Appendix 3). Inclusions continued until the interviews perceived data saturation.

Interview recordings were transcribed verbatim (Amir Haourara, Florent Lidec, Catherine Boivin, and RBJ) and then coded (Anis Harbi, CB, GR, MD, Olivia Rousset Torrente, RBJ, and SF) to facilitate thematic analysis using a General Inductive Approach using the methodology developed by Thomas [16]. Triangulation coding was conducted with open-source Sonal software (Alex Alber, Université F. Rabelais [Tours]). The coding process was developed over time following several meetings among the research team (CB, GR, MD, and SF) and then analyzed (GR).

Ethics Approval

The study was approved by the Inserm Ethics Independent Committee (00003835, protocol 2016/43NI) and then registered with French data protection authority (2008669). Verbal consent was obtained from each participant prior to interviews.

Results

Sociodemographic Characteristics

Participants included 11 doctors and 9 nurses who conducted medical examinations at OFII. In all, 14 (70%) participants were female. Their median age was 58 (range 25-67) years, with a median of 25 (range 2.5-40) years of professional experience and a median of 3 years (range 1 month to 22 years) of working at OFII. Combined, the participants conducted medical examinations in 9 languages. Only 1 (5%) participant spoke solely French. Details are displayed in Table 1.
Table 1. Health professionals’ sociodemographic characteristics (N=20).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>6 (30)</td>
</tr>
<tr>
<td>Female</td>
<td>14 (70)</td>
</tr>
<tr>
<td><strong>Age group (years)</strong></td>
<td></td>
</tr>
<tr>
<td>20-29</td>
<td>1 (5)</td>
</tr>
<tr>
<td>30-39</td>
<td>1 (5)</td>
</tr>
<tr>
<td>40-49</td>
<td>3 (15)</td>
</tr>
<tr>
<td>50-59</td>
<td>9 (45)</td>
</tr>
<tr>
<td>60-69</td>
<td>5 (25)</td>
</tr>
<tr>
<td><strong>Localization</strong></td>
<td></td>
</tr>
<tr>
<td>Center 1</td>
<td>6 (30)</td>
</tr>
<tr>
<td>Center 2</td>
<td>5 (25)</td>
</tr>
<tr>
<td>Center 3</td>
<td>8 (40)</td>
</tr>
<tr>
<td>Center 4</td>
<td>1 (5)</td>
</tr>
<tr>
<td><strong>Position</strong></td>
<td></td>
</tr>
<tr>
<td>Medical doctor</td>
<td>11 (55)</td>
</tr>
<tr>
<td>Nurse</td>
<td>9 (45)</td>
</tr>
<tr>
<td>Having a health care speciality&lt;sup&gt;a&lt;/sup&gt;</td>
<td>6 (30)</td>
</tr>
<tr>
<td><strong>Seniority as a health professional (years)&lt;sup&gt;a&lt;/sup&gt;</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;10</td>
<td>1 (5)</td>
</tr>
<tr>
<td>10-19</td>
<td>6 (30)</td>
</tr>
<tr>
<td>20-29</td>
<td>4 (20)</td>
</tr>
<tr>
<td>30-39</td>
<td>7 (35)</td>
</tr>
<tr>
<td>40-49</td>
<td>1 (1)</td>
</tr>
<tr>
<td><strong>Seniority at OFII&lt;sup&gt;b&lt;/sup&gt; (years)&lt;sup&gt;a&lt;/sup&gt;</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;1</td>
<td>3 (15)</td>
</tr>
<tr>
<td>1-9</td>
<td>9 (45)</td>
</tr>
<tr>
<td>10-19</td>
<td>4 (20)</td>
</tr>
<tr>
<td>≥20</td>
<td>2 (10)</td>
</tr>
<tr>
<td>Having another job outside of OFII job&lt;sup&gt;a&lt;/sup&gt;</td>
<td>16 (80)</td>
</tr>
<tr>
<td><strong>Previous professional experience (before OFII)&lt;sup&gt;b&lt;/sup&gt;</strong></td>
<td></td>
</tr>
<tr>
<td>With migrants</td>
<td>10 (50)</td>
</tr>
<tr>
<td>HIV, HBV&lt;sup&gt;c&lt;/sup&gt;, or HCV&lt;sup&gt;d&lt;/sup&gt; prevention activities</td>
<td>8 (40)</td>
</tr>
<tr>
<td><strong>Native language&lt;sup&gt;a&lt;/sup&gt;</strong></td>
<td></td>
</tr>
<tr>
<td>French</td>
<td>15 (75)</td>
</tr>
<tr>
<td>Not French</td>
<td>5 (25)</td>
</tr>
<tr>
<td><strong>Number of foreign languages spoken&lt;sup&gt;a&lt;/sup&gt;</strong></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1 (5)</td>
</tr>
<tr>
<td>1</td>
<td>7 (35)</td>
</tr>
<tr>
<td>2</td>
<td>9 (45)</td>
</tr>
<tr>
<td>Characteristic</td>
<td>Participants, n (%)</td>
</tr>
<tr>
<td>---------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>3</td>
<td>3 (15)</td>
</tr>
</tbody>
</table>

*Missing data for 1 participant.

b OFII: Office Français de l'Immigration et de l'Intégration (French Office of Immigration and Integration).
c HBV: hepatitis B virus.
d HCV: hepatitis C virus.

**Thematic Analysis**

In all, 5 major themes and 11 subthemes were defined during the thematic analysis, detailed in the following diagram (Figure 1).

**Figure 1.** Thematic tree. HP: health professionals; LFP: limited French proficiency; OFII: Office Français de l'Immigration et de l'Intégration (French Office of Immigration and Integration); TROD: test rapide d'orientation diagnostique (rapid diagnostic test).

**Positive and Negative Attitudes Toward an App**

All HPs reported having used (at least once, in their private life) a translation app such as Google Translate (the most cited). Most of them used one at OFII, in absence of a better solution, by only formulating wording-simplified close-ended questions. Some noticed mistranslations, which caused no overall misunderstanding though. They reported that migrants with LFP spontaneously used translation apps too. HPs were not opposed to a tool that would help them do their job that would also benefit their patients, “I’m in favour [of a tool] because it is for the patient’s benefit” (male doctor). One doctor thought that an app could be used by doctors to “give their opinion and explain to the patient” what our objective is [as an HP] in the context of rapid screening” to better communicate with the patient (male doctor).

However, some HPs who were concerned about the effectiveness of an app to actually improve the uptake of screening because of the migrant with LFP’s preconceived notions of the screening process. “Not understanding the added value” or “a lack of trust” no matter how good the app is were mentioned. “For [migrants with LFP], there is no interest. They have already got it into their heads ‘I don't want to do it,’ or else ‘I’m going to do it,’ but it’s not the health benefit that motivates them” (female nurse).

In addition, it was suggested that a new tool would be difficult to implement “because people are on the move, left and right during the medical consultation at OFII”, uh...I don’t know
when they would have time to use it” (female nurse). It was also expressed that instead of a new tool, it would be better to revise current screening practices (female nurse).

What Type of Tools?

Printed Material–Improvement of What Exists

One participant suggested that when planning a new tool, it should begin with revising the existing texts that are used: “translate our proposition [to conduct a rapid test]. To have that information already translated for the LFP…it’s super important” (female nurse).

Currently, HPs at OFII reference a binder containing documents translated into 13 languages, with standard sentences used to ask sociodemographic and health history questions. Several participants were satisfied with a paper-based system. One participant, who relies on this printed material, said that he preferred “printed documents.”

For me, a good tool is a printed document [translated] into the patient’s mother tongue...For me, I get out the document...there you go. It’s all I need. The LFP reads it and...while doing so, I try to follow what they are reading because I know [what’s written]. Basically, I know what the questions correspond to...they answer, and then it saves me from having to rely on Google translate or calling an interpreter. It’s quick.” [male doctor]

Some centers have created their own documents so that additional languages can be offered; therefore, practices differ from center to center. Another participant said that a new tool could simply be an improvement to the existing paper-based system. The new tool could therefore be “a sheet of paper” (female doctor) or “a poster” (female nurse) placed in the waiting room, since “some [migrants] have long wait times” (female nurse).

Other participants are resistant to the idea of the new tool having a paper format because it assumes that the patient is literate, as two participants explained: “Some can’t even read, so...there’s a disconnect” (female nurse).

Modes of Electronic Presentation

Informants spoke of how a new app might augment existing technology that they already have the habit of using. The new tool could be used “over the phone” (female doctor) or “on a tablet” (female nurse), as well as “A tool on the internet, on the computer” (male doctor), such as “Google Translate” (male doctor). Another mentioned “artificial intelligence and technology” (male doctor).

Functionalities

A Precise, Reliable, and Instantaneous Translator With Audio

Participants wanted the device to be precise and reliable, “the formulation must be really refined” (male doctor), because the HPs need a “support that is reliable” (female nurse) and trustworthy.

The main expectation of HPs is that a new device would be a “translator” (male doctor). HPs would appreciate a translator that has an audio function and works simultaneously with their speech. This would make it possible to solve the illiteracy of certain migrants: “We would speak, and [the tool] would translate at the same time, for people who can and who cannot read, there would be the audio” (female nurse). This would guarantee the confidentiality of information (compared to a third party, such as a professional or informal interpreter):

You just put your language in, and then when you are speaking, it translates immediately, the person understands, they speak...Well, I will answer you frankly. The ideal for me is simultaneous translation, perfect. There you go, if you want efficiency for work and confidentiality during the consultation, that’s all. [male doctor]

Easy to Use and Understand

HPs imagined a simple tool that is both easy to use and understand: “simple words. It shouldn’t ask too many questions. The [migrant] must also be able to understand...it has to be easy to use, practical for the consultation.” (male doctor). Another participant wanted the tool to translate “according to the levels” of knowledge of the migrant with LFP (female nurse).

Modalities of Information Provision

Diagrams, Pictograms, and Images

HPs suggested adding “small diagrams” (female nurse): “why not include images?” (male doctor). Visual communication was seen as “a supplement; it can...help” (female doctor). One participant stressed that “it’s about the drawings” (female nurse). Another recalled that “there are people who have an educational level which is not [enough to read]...there are quite a few, [so] yeah, [there’s a need] for pictograms, drawings” (male doctor). However, one expressed doubt about images having divergent cultural meanings: “a different interpretation of a pictogram, they [the migrants] are not always perceived in the same way” (female doctor). Although images were considered to be useful, there was concern that they were polysemous.

Video

A video was considered to be a more effective way to transmit a message compared to plain text, audio, or pictograms: “it would be the most effective” (female nurse). Several participants would like the same characteristic: a “little video” (female nurse) of “short duration” (female doctor), nothing longer than “3 minutes” (female nurse).

However, one participant doubted the added value of video: “In respect to a video, when you are infected, you will already have a document to read...would an image or video add anything in addition to the text? Hmm...” (male doctor). Another participant expressed the potential difficulty of using video in an OFII waiting room: “a video, in my opinion, would be the most informative. But in fact, putting that in place, I’m not sure it’s very easy” (female nurse). Therefore, a short video was considered the best way to present information but challenging to implement.

https://humanfactors.jmir.org/2023/1/e41861 | JMIR Hum Factors 2023 | vol. 10 | e41861 | p.84
(page number not for citation purposes)
Audio, Another Solution to Illiteracy

The use of “a voice” or audio was mentioned by 2 participants: “It’s not bad [the audio], like an exhibition [at a museum]” (female nurse). Another participant saw the benefit as a means to overcome illiteracy: “Perhaps with audio, yeah, because we still have the issue with people who can’t read” (female nurse).

Anticipated Outcomes

To Obtain Medical Information

Several doctors mentioned numerous questions that they would like to see in the tool. These were questions that they need to pose concerning vaccinations, surgical operations, tattoos, previous and frequent illnesses, procreation, and risks taken—similar to “a classic medical exam” (male doctor).

To Adapt to the Patient’s Needs

The tool was also envisioned by HPs to extend their work beyond the screening process. One participant envisaged a tool that could provide immediate feedback during the consultation that could be personalized according to “the person in front of us, if he says to us: ‘I have been operated on,’ it [the app] informs us, and we easily understand” (male doctor), enabling the HPs to be able to conduct a more thorough risk assessment. Several HPs had noticed a lack of knowledge about the medical interventions that the patients had experienced and wanted the new tool to adapt to this lack of knowledge:

If the person says, “I had surgery. Well, they cut my stomach open.” “What exactly did they have done?” There are people who don’t understand what kind of interventions they’ve had...It seems useful to me [to inform people] of the basic things at least, to provide information. [male doctor]

Another participant took up the same theme by proposing that the tool help the HP educate the migrant with LFP: “I could explain the mode of transmission!” (female nurse). Another participant suggested that the tool encourages migrant with LFP to educate themselves with the HPs they will meet in consultations: “If you want more information, the person in front of you can help you” (female nurse).

To Facilitate the Flow of Information Between the HP and Migrant With LFP to Offer the TROD

The tool was described as a potential aid to the medical relationship in that it could be used before the migrant is asked about the TROD, it would say to the migrant: “‘Here, we are looking for such things, and here is what we offer you’...then they know [what] we are looking for” (male doctor). It would “give them [the migrant] confidence from the start [of the migrant’s arrival at OFII].” One participant expected “that [the tool] would not disturb the climate of the medical consultation” (male doctor), “if there is something specific, well, we would ask questions” (male doctor).

To Measure and Target Levels of Health Literacy

Several participants talked about a new tool offering opportunities for data collection that were not possible within the current protocol. Principally, this was the inclusion of a health literacy test that could be integrated into the tool. Participants reported that “it might be interesting for us to know” the migrant’s level of understanding (female nurse). This test would contain questions such as “Do you know these different diseases, hepatitis B, hepatitis C, HIV? Do you know how they are transmitted?” (male doctor). The tool would then report the migrant’s answer to the HPs who would be able to adapt their presentation to the patients’ level of understanding.

Discussion

Principal Findings

Despite doubts of a digital communication tool and ingrained habits using printed translated materials, participants favored the integration of an innovative digital tool to enhance communication with migrants with LFP in the context of rapid screening. Although some participants found current printed materials effective, this communication method is unidirectional and thus does not promote a dialogue between the patient and provider. Furthermore, HPs spoke of the perceived benefit of using translated and culturally adapted multimedia content to better communicate with their patients and enhance the consultation experience for all parties, thus requiring a digital solution. These positive expectations will attract HPs to the innovative tool once developed and will play a crucial role in the mobilization of resources for its successful implementation [17].

A systematic review investigating the use of electronic tools to help increase testing in migrants with LFP, conducted by our research team, found that translation apps provide better communication with HPs and have a high acceptability of use [10]. In terms of a new tool, our research found that HPs spoke most frequently of an easy to use (Figure 1, subtheme 3.2), accessible app with multiple features, including visual (Figure 1, subthemes 4.1-4.2) and audio (Figure 1, subtheme 4.3) components along with an accurate, reliable, and instantaneous translation (Figure 1, subtheme 3.1); data collection; adaptive content; and interpretation functions (Figure 1, subthemes 5.1, 5.2, 5.3, and 5.4).

A cross-sectional study comparing patient-provider communication with IT-mediated communication versus face-to-face communication found the same level of effectiveness, although patients prefer face-to-face communication with their provider [18]. Although we did not study the patients’ perspective, participants expressed interest in a new tool that could be tailored to the medical consultation (Figure 1, subtheme 5.2) and adapted to the patient’s needs. The patients’ perspective, however, needs to be further explored to understand their perceptions of mixed communication methods, which would include digital and face-to-face communication during the same consultation.

During the interviews, HPs spoke of their need for a tool that could obtain medical information from their patients, facilitate communication to offer the TROD, and measure patients’ level of health literacy, which would then provide educational material to patients. A systematic review found that touchscreen apps could help patients with limited health literacy better understand medical information and provide education on medical
treatments [19]. Furthermore, a prototype to support patient-provider interaction in chronic HIV care found that patients want an app that is easy to use and intuitive while meeting confidentiality and security standards [20].

The use of artificial intelligence can provide user-targeted messages to increase the effectiveness of communication and education [21]. Such technology can reach wider and often harder-to-reach audiences than traditional means of communication. One artificial intelligence chatbot deployed in India to encourage conversations on sexual and reproductive health found that the app was an educationally beneficial tool for reaching vulnerable audiences [22].

An app, in the patient’s native language, could help educate patients about the benefits of being screened while at the same time helping the HPs to propose and conduct rapid screening. This would increase the patient’s understanding of HIV, HBV, and HCV including modes of transmission and risk reduction practices. Interfacing with an app could also create a more comfortable context to learn about topics such as sex and high-risk situations than if the HPs interview them on these subjects.

With the increase in mobile technologies in the health sector, an app would be an innovative mobile health approach to increase the screening rate of HIV, HBV, and HCV in an effort to achieve national and international objectives.

**Study Strengths**

This study is the first of its kind in the French context. Interviews with both nurses and physicians who conduct medical exams and rapid screening tests at different centers allowed us to gain a better overall understanding of how language barriers effect medical consultations. It also provided us with insight into what HPs want from a communication tool to overcome language and cultural barriers. We found that a better-adapted communication intervention could help HPs overcome language barriers with migrants with LFP and ultimately, increase screening rates.

**Limitations**

A limitation of our study is that this research was conducted in only one context and therefore not representative of migrant screening throughout France. Although we included centers in both the Paris region and in other areas of France, there are more than 30 OFII centers, and immigration is not homogenous throughout France.

**Conclusion**

Our research allows us to better understand the expectations of health care providers for new technological solutions. These expectations are crucial to the development and adoption of the technology. We have explored the potential format, characteristics, functions, content, and use of a new technology to communicate with migrants with LFP. In terms of an app, we found positive expectations and support from HPs to develop and use an app in the patient-provider relationship to overcome language and cultural barriers.

This information will be used to develop an app and implement the ApiDé study [23], which aims to validate a communication app in an attempt to address language barriers and, ultimately, increase screening rates of migrants with LFP in France.

**Acknowledgments**

We would like to thank the administration and health professionals from the French Office of Immigration and Integration (Office Français de l’Immigration et de l’Intégration [OFII]) for their collaboration and participation in the interviews. A special thanks to the Professional Quality of Life (PROQOL) research group for their administrative support and assistance in validating the interview guide and the STRADA pilot committee for their continued scientific support. We also thank Dr Thanh Le Luong from OFII for her help in facilitating the organizations of interviews.

We would like to give thanks to the transcription team not listed as authors: Amir Haouara, Catherine Boivin, and Florent Lidec. We also thank the triangulation coding team not listed as authors: Anis Harbi and Olivia Rousset Torrente, and for their time coding interviews.

Funding for this project was received from the Asylum, Migration and Integration Fund (AMIF; award FAMI-I-16-342); the OFII; ViiV Healthcare (grant CT180187); and Gilead.

Neither AMIF, ViiV, nor Gilead had a role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

The scientific committee consisted of David Zucman, France Lert, Françoise Roudot-Thoraval, Karine Lacombe, Florence Nicolaï Guerbe, Philippe Sogni, Alain Sobel, and Svetlana Dimi.

**Authors’ Contributions**

MD contributed to conceptualization, supervision, and project administration. SF, MD, and CB contributed to methodology. GR contributed to software and data curation. SF and MD contributed to validation. GR and CB contributed to formal analysis. SF contributed to investigation and resources. CB wrote the original draft. JC, GR, MD, and OC contributed to writing—review and editing. GR and CB contributed to visualization. MD and OC contributed to funding acquisition.
Conflicts of Interest
None declared.

Multimedia Appendix 1
French interview guide.
[DOCX File, 21 KB - humanfactors_v10i1e41861_app1.docx ]

Multimedia Appendix 2
English interview guide.
[DOCX File, 19 KB - humanfactors_v10i1e41861_app2.docx ]

Multimedia Appendix 3
Consolidated criteria for reporting qualitative research (COREQ) guideline checklist.
[PDF File (Adobe PDF File), 414 KB - humanfactors_v10i1e41861_app3.pdf ]

References


Abbreviations

COREQ: Consolidated criteria for reporting qualitative research
HBV: hepatitis B virus
HCV: hepatitis C virus
HP: health professional
LFP: limited French proficiency
OFII: Office Français de l’Immigration et de l’Intégration (French Office of Immigration and Integration)
TROD: test rapide d’orientation diagnostique (rapid diagnostic test)

Edited by A Kushniruk; submitted 16.08.22; peer-reviewed by M Kapsetaki, M Metcalf; comments to author 04.10.22; revised version received 28.11.22; accepted 11.12.22; published 03.02.23.

Please cite as:
Brown C, Roucoux G, Dimi S, Fahmi S, Jeewan RB, Chassany O, Chaplin J, Duracinsky M
Evaluating Clinician Expectations of mHealth Solutions to Increase Rapid-Screening for HIV and Hepatitis in Migrant Populations in France: Qualitative Study
JMIR Hum Factors 2023;10:e41861
URL: https://humanfactors.jmir.org/2023/1/e41861
doi: 10.2196/41861
PMID: 36735323

©Carter Brown, Guillaume Roucoux, Svetlane Dimi, Saleh Fahmi, Raj-Banou Jeewan, Olivier Chassany, John Chaplin, Martin Duracinsky. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 03.02.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

https://humanfactors.jmir.org/2023/1/e41861 JMIR Hum Factors 2023 | vol. 10 | e41861 | p.88
Acceptability of a Health Care App With 3 User Interfaces for Older Adults and Their Caregivers: Design and Evaluation Study

Joo Chan Kim¹, MSc; Saguna Saguna¹, BIS (Hons), MIT, PhD; Christer Åhlund¹, LicEng, MSc, PhD
Division of Computer Science, Department of Computer Science, Electrical and Space Engineering, Luleå University of Technology, Skellefteå, Sweden

Corresponding Author:
Joo Chan Kim, MSc
Division of Computer Science
Department of Computer Science, Electrical and Space Engineering
Luleå University of Technology
Forskargatan 1
Skellefteå, 931 77
Sweden
Phone: 46 722254785
Email: joo.chan.kim@ltu.se

Abstract

Background: The older population needs solutions for independent living and reducing the burden on caregivers while maintaining the quality and dignity of life.

Objective: The aim of this study was to design, develop, and evaluate an older adult health care app that supports trained caregivers (ie, formal caregivers) and relatives (ie, informal caregivers). We aimed to identify the factors that affect user acceptance of interfaces depending on the user’s role.

Methods: We designed and developed an app with 3 user interfaces that enable remote sensing of an older adult’s daily activities and behaviors. We conducted user evaluations (N=25) with older adults and their formal and informal caregivers to obtain an overall impression of the health care monitoring app in terms of user experience and usability. In our design study, the participants had firsthand experience with our app, followed by a questionnaire and individual interview to express their opinions on the app. Through the interview, we also identified their views on each user interface and interaction modality to identify the relationship between the user’s role and their acceptance of a particular interface. The questionnaire answers were statistically analyzed, and we coded the interview answers based on keywords related to a participant’s experience, for example, ease of use and usefulness.

Results: We obtained overall positive results in the user evaluation of our app regarding key aspects such as efficiency, perspicuity, dependability, stimulation, and novelty, with an average between 1.74 (SD 1.02) and 2.18 (SD 0.93) on a scale of −3.0 to 3.0. The overall impression of our app was favorable, and we identified that “simple” and “intuitive” were the main factors affecting older adults’ and caregivers’ preference for the user interface and interaction modality. We also identified a positive user acceptance of the use of augmented reality by 91% (10/11) of the older adults to share information with their formal and informal caregivers.

Conclusions: To address the need for a study to evaluate the user experience and user acceptance by older adults as well as both formal and informal caregivers regarding the user interfaces with multimodal interaction in the context of health monitoring, we designed, developed, and conducted user evaluations with the target user groups. Our results through this design study show important implications for designing future health monitoring apps with multiple interaction modalities and intuitive user interfaces in the older adult health care domain.

(JMIR Hum Factors 2023;10:e42145)  doi:10.2196/42145

KEYWORDS

Internet of Things; health monitoring; older adults; augmented reality; user experience; independent living; design study; mobile phone
Introduction

Background

According to a United Nations report, the number of people aged ≥65 years in 2020 was approximately 727 million, which is expected to increase to 1.5 billion by 2050 [1]. As the proportion of older adults increases, the demand for older adult care services increases [2,3]. In particular, older adults who live independently require care because of physical and mental health vulnerabilities such as physical constraints, poverty, loneliness, and depression [1]. Their relatives who live independently have difficulty visiting them every day because of distance and time issues. Therefore, older adult care services to improve life satisfaction are necessary. However, the burden on caregivers and relatives keeps growing owing to the aging and increase in the older adult population. According to the American Association of Retired Persons and National Alliance for Caregiving report in 2020 [3], 18% of caregivers covered multiple people in 2015. This ratio increased by 6% over 5 years to 24%. In addition, 54% of caregivers were aged >50 years, and 21% of family caregivers (i.e., relatives) reported that caregiving had worsened their health. This phenomenon has worsened because of the pandemic [4,5]. In this context, the necessity of assistive technology to support relatives and caregivers in reducing their burden has continuously grown.

To support caregivers and relatives, the latest status information of an older adult can be provided by an Internet of Things (IoT)-based system. IoT is a technology widely used for collecting data about a person and their environment to enable the system to understand the information of their context.

For example, a sensor attached to a human body could work as a heart rate monitor [6], and in another case, a sensor can read air quality pollutants to work as an air quality sensor [7]. As data need an interface to be delivered to a user, efficient data delivery is as essential as data collection. Augmented reality (AR) draws interest from researchers as a technology that could enhance user engagement [8] and enrich data presentation for better accessibility [9]. The properties of both technologies are attractive; hence, research to improve the merits of IoT and AR has been conducted by combining them since those 2 technologies gained attention [10-12].

There is a need for more research on IoT platform–based AR apps, especially regarding users’ perception of an app’s user interfaces (UIs) and acceptance of the technologies used in the context of health monitoring of an older adult by caregivers and relatives. For example, in the Internet of Things within health and care (iVO) project [13], older adults’ activities are sensed by IoT devices, and anomalous events are reported to their relatives via SMS text message [14]. However, efficiently conveying comprehensive information about an older adult’s state to their relatives and caregivers could be done with well-designed UIs rather than SMS text messaging services. In this case, the user experience of the app and user acceptance of the app’s UI with the technologies used should be analyzed based on the user’s role to understand the effectiveness factor. Furthermore, AR is useful for visualizing data. Hadj Sassi and Chaari Fourati [15] showed that displaying real-world data on a 3D AR map identical to the real world is beneficial in understanding the data. However, their user evaluation focused on UIs’ usability related to performing a task and generic user experience. On top of the user experience evaluation, an in-depth analysis of user acceptance regarding interaction modalities and data presentation designs depending on the users’ characteristics (e.g., age, gender, and experience) is needed.

Regarding user acceptance, the definition varies based on the purpose of use [16]. Technology acceptance model, developed by Davis [17] is a widely used approach to measure acceptance [16], and it proposes that user acceptance is determined by 2 factors: “perceived usefulness” and “perceived ease of use” [17]. “Perceived usefulness” means a user’s perception of the technology, whether it is helpful for their task. “Perceived ease of use” is a user’s feeling of how easy it is to use the technology. These 2 factors influence a user’s belief about the technology, which determines acceptance and use [17]. In addition, a user’s characteristics, such as age, gender, experience, and voluntariness of use, can affect their perception of the technology and, hence, influence user acceptance [18]. In our study, we used the definition of user acceptance by Dillon and Morris [19]: “the demonstrable willingness within a user group to employ information technology for the tasks it is designed to support.” On the basis of this definition, we examined the reason for preference by users in terms of “perceived usefulness” and “perceived ease of use” depending on their characteristics, especially on their role (e.g., caregiver, relative, and older adult), from user evaluations to demonstrate a user’s willingness and, thus, the user acceptance regarding interaction modalities and data presentation designs.

Objectives

Our design study aimed to conduct user evaluations on both the app and its 3 different UIs designed for caregivers, relatives, and older adults to identify the app’s user experience and the factors that affect user acceptance of each UI depending on their characteristics, especially on the participants’ role. In this study, we grouped caregivers and relatives into 1 category, “caregivers,” and separated them based on whether they had training experience in health care services as experts. Accordingly, relatives were labeled as “informal caregivers,” whereas other trained experts were grouped as “formal caregivers.” By understanding the relationship between a user’s role and UI, we can adapt the UI designs to efficiently inform an older adult’s state. Each UI has a different concept. For example, one UI is designed on a tile-based template, whereas another UI uses a 3D map to present data within its virtual space. The last UI overlays AR contents around a user’s face, and the data are delivered through AR contents. Although the 3 UIs have distinctive design themes, the data displayed on every UI are almost identical, and the interaction modalities supported on each UI are similar with minor differences. On the basis of the meeting with iVO project participants, we hypothesized that informal caregivers would prefer the tile-based UI with audio-based interaction (e.g., voice command inputs and audio outputs) because of the simplicity of data presentation and hands-free property. In contrast, formal caregivers would prefer the map-based UI with touching and reading capabilities because of the different data levels, intuitive data visualizations, and ease of navigation.
of use while visiting an older adult’s residence. Meanwhile, we assumed that older adults would prefer the AR-based UI with touching and reading capabilities because of engagement, intuitiveness, and easiness. To consider a practical use case that requires mobility, we implemented and evaluated these 3 UIs on mobile devices. We designed our app to be able to use an IoT platform to receive an older adult’s daily activity data.

In this design study, we made the following contributions in the context of health monitoring of an older adult: (1) we designed and implemented the app with 3 initiative UIs for formal and informal caregivers to support the care of older adults using IoT; (2) we conducted user evaluations to analyze user experience and user acceptance of the app and its UIs to identify the relationship between the user’s role and their acceptance of a particular UI, and this would emphasize the necessity for diversity in interaction modalities and UIs; and (3) we observed overall positive user acceptance of using AR by participants and especially among the older adult participants, along with ideas on how AR can be used further in the context of older adult health care.

The design of our system and app is described in detail in the following section. Next, we describe the user evaluation procedure and the data analysis. Then, we present the results of the data analysis, categorized as overall impression and user acceptance, to show participants’ impressions of our app. This paper ends with a discussion of our design study’s implications for the health care monitoring domain and its contributions to future studies.

Methods
Development
This section explains the system environment that was used to collect human behavioral data in people’s residences. We then present the design process to build the UIs along with the target device for running our app.

App Environment
In this study, we used the IoT platform Societal Development Through Secure IoT and Open Data for monitoring a person’s daily activities, as shown in Figure 1.

Figure 1 shows the IoT platform being used where arbitrary sensors can be connected; data are gathered, stored, and processed to identify activity in homes. The service designed, developed, and evaluated is the older adult well-being service in Figure 1. Shahid et al [14] give more details on data processing and analytics that designed a framework for preprocessing and processing the data and activity recognition models based on data from the off-the-shelf sensors and IoT devices installed in homes to learn daily patterns of different activities and detect anomalies.

This study aimed to evaluate participants’ user experience and impression of the UIs. The work done in the iVO project [13] also forms the basis for the need to design our app and its UIs as, in that study, information was shared via SMS text message notifications. However, during repeated interviews and communications with the participants and their caregivers, a need for an app with more detailed information to view was observed. The primary data used for visualization in this study were (1) duration of being active (ie, activeness), (2) duration of being still (ie, stillness), (3) duration of staying in a room (ie, as both active and still), and (4) transition logs from one room

Figure 1. Framework for Societal Development Through Secure Internet of Things and Open Data (SsIO) health care services. IoT: Internet of Things.
to others. In addition, Textbox 1 describes all the data used to detect activity in each room.

**Figure 2** illustrates an example of IoT sensor installations in an older adult’s residence. We installed nonintrusive IoT sensors in each room, and the actual sensor installations were adjusted to the room design and available appliances in the older adults’ houses. The app designed for this study could be used to check for both normal and abnormal activities.

The real-world behavioral data collected through the iVO project [13] were used to create a generic older adult’s 3-day behavioral activity pattern and used as sample data for our app.

Textbox 1. Example of collected and processed data for abnormal activity detection.

- Bathroom
  - Duration of stay and number of visits during sleeping time
- Bedroom
  - Duration of sleep
- Living room
  - Duration of stay and television use
- Kitchen
  - Duration of stay and number of appliances used
- Balcony
  - Nighttime visits

**Figure 2.** Sensor installations in an older adult’s residence for collecting daily activity data. IoT: Internet of Things.

**App Design**

**Overview**

On the basis of the design principles [20-22], we designed a prototype app that consisted of 3 UIs through several iterations. As we wanted to identify various useful design elements and collect diverse feedback regarding UI design, we prepared 3 different UIs with unique concepts. Once the core features of each UI were implemented, such as tiles with large icons and text, a 3D rotatable map, and floating AR contents around a user’s face, we performed user tests with our colleagues to identify possible improvements in user experience and usability perspectives. We updated the visibility and readability of
information on each UI, including an aligned menu design for better accessibility with an intuitive navigation procedure. The following sections describe the interaction modalities and UIs used in the user evaluations.

**Design Principles**

The target user group in our study included older adults in particular; thus, the UI design should consider the age-related elements that could affect the user experience [20-22], for example, big font size and high graphic clarity for visual elements; low-frequency perception and additional stimulation, such as a vibration of a mobile device, for auditory interaction; and a rule-based color theme and simplified menu navigation for cognitive processes. We used these elements as fundamental design principles for our 3 UIs.

For one of the UIs, we used a tile-based design along with text and pictograms inspired by commercial apps such as the Oura Ring [23] and Apple’s home app [24]. We expected that the strength of the tile-based design would be the simplified data presentation with intuitiveness.

In the map-based UI, we used a 3D map and several graphical elements on that map to present information. The information presented in all 3 UIs (Figure 3), including the AR-based UI, was similar; however, we found that data presented on a virtual map that refers to a real-world space could further improve the intuitiveness of the information [15,25].

According to our literature survey, the properties of AR have positive effects such as motivation and intuitive data visualization [12,15,26]; therefore, we decided to use these effects in our app to support older adults. Although some studies in the health care domain relied on a printed marker [9,15], we decided to use a face filter style of UI for AR in the older adult health monitoring service to evaluate the acceptance level of AR by older adults.

**Interaction Modality**

We used multimodal interaction to provide flexibility in the UI for older adults [20,21]. However, simply increasing the number of interaction modalities poses a potential failure to achieve effective multimodal interaction [27,28]. Therefore, we used basic interaction modalities that modern mobile devices support instead of adding more modalities using external devices. For example, we enabled touch and voice command and used facial parts as visual cues for input modalities. In contrast, visual elements, sound, and a device’s vibration were used as visual, audio, and haptic output, respectively. The mobile device vibrated when the user pressed a button that contained an abnormal event or an animation to play. Therefore, the vibration was an additional modality to emphasize a notification rather than the principal channel for delivering information, such as visual elements and sounds. We enabled every interaction modality in all UIs as we wanted to evaluate the end users’ initial impression of our app and its UIs that were similar to the final product.

**UI Design**

We designed 3 UIs that present similar information but in different forms. Our primary UI design principle was to achieve a proper level of intuitiveness for reducing the cognitive process of finding and understanding the information. The reason for choosing design principles was that our app’s target user group was older adults, for whom the intuitiveness of data presentation is an essential factor. Figure 3 illustrates our 3 UIs to aid health care tasks performed by formal and informal caregivers of older adults.

We referred to several findings from related studies regarding the UI design for older adults in our app’s UI design, for example, a large font, button, and image size for better visibility; consistent color scheme to increase the readability of the information; simplified menu navigation for ease of use and fast data access; support for offline accessibility to prevent user experience interruption; and simplification of data visualization for intuitive information delivery [20,22,29].

---

**Figure 3.** The 3 user interfaces display similar information in different formats. AR: augmented reality.
To ensure consistency between UIs, all 3 UIs have 5 shared features marked in the tile view in Figure 3. First, the top 3 buttons are for setting the window to select data. The date button is used for choosing the date. The time button is used for selecting the time window. The room button decides which room data the user wants to see. Second, the speaker button is used to play the audio for reading out the information. When the audio is playing, pressing this button stops the audio. Third, switching the language between English and Swedish is done by pressing the language button. Fourth, the microphone button enables the voice command feature for interacting with the app using a human voice. The voice command consists of 3 keywords to correctly configure the system for receiving data: date, time, and room name. Finally, the bottom 3 buttons are for switching between UIs.

Each UI has a unique design concept for presenting information to users in addition to these common features. We designed the tile view and the map view to provide as much data as possible, from overview to detail, to formal and informal caregivers. In contrast, the AR view was designed for older adults. We decided to present minimum data in the AR view based on interviews with older adults [14,30]. We found that older adults tend not to show interest in the detailed report of their daily activity; therefore, we simplified both the level of data and the visualization complexity.

First, the tile-based UI that uses rounded squares with large icons with a minimum amount of text to describe the information is named tile view (Figure 3). When abnormal behavior is detected, a correlated tile displays the exclamation mark icon to emphasize that the user has to be aware of it. Each tile is clickable, and the information regarding the pressed tile is played as an audio output. In addition, the device vibrates when the tile with an exclamation mark icon is pressed. A transition log from the selected room to others is presented when a specific room is selected, such as in the tile view in Figure 4.

**Figure 4.** The user interfaces display a transition log between rooms in the tile view and map view. In the augmented reality (AR) view, a feature captures the screen image and AR information for sharing with other users.

Second, the 3D map–based UI presenting a person’s behavior data on a 3D-modeled floor plan is named map view (Figure 3). The 3D map is an actual floor plan of the user’s residence, thereby expected to improve the UI’s intuitiveness. The circle icon with a progress bar indicates the percentage ratio of activeness and stillness of a person in each room. The cylinder in a room also represents the activeness and stillness of a person through the cylinder’s height. When an event such as kitchen appliance use or abnormal behavior occurs, additional icons are visible next to the circle icon. For example, when the coffee pot is used during lunchtime, a coffee pot icon is displayed. If not, an exclamation mark icon is visible to represent that abnormal behavior is detected. The map can be rotated by dragging it with a finger, and the view on the map is changeable from perspective view to top view and vice versa. The transition log ordered by time is listed below the map. A correlated trajectory line on the map is animated to highlight the information when the user clicks on one of the buttons on the log list. In addition, the information related to the selected log is played as an audio while the trajectory line is animated. The buttons on the log list can have an exclamation mark icon when the log contains abnormal behavior. In this case, the device vibrates once the user presses the button. The transition log is also provided in another panel depicted in the map view in Figure 4 when the user selects the circle icon on the map. An additional pop-up window appears to show detailed information about abnormal behavior when the user clicks on the exclamation mark icon on this panel.

Last, we used the ARCore (Google) face-tracking feature [31] to use the user’s face as a marker for AR (see the AR view in Figure 3). The data are presented as AR text with AR icons floating around the user’s face. As a result, the user does not need to prepare a printed marker to visualize AR objects. The AR icons are clickable. Once the AR icon is pressed, correlated information is played through audio, and the device vibrates as well. Although the tile view and map view are designed for both
formal and informal caregivers, the AR view is designed for older adults. We foresaw that older adults could accept the AR view for the following reasons. First, we minimized the information given in the AR view by focusing on the main activities in each room. Second, AR objects would make older adults engage in using the AR view. Third, the user could capture an image of their face along with data visualized through AR objects. The AR view in Figure 4 shows the captured image with data as AR objects. This captured image could be shared with formal and informal caregivers to inform of the user’s latest state.

The flowchart and user flow of each UI are provided in Multimedia Appendix 1 and Multimedia Appendix 2, respectively. In addition, the summary of each UI’s details, including target user, interaction modalities, and unique features, is presented in Multimedia Appendix 3.

Target Device and Configuration
We used Samsung Galaxy Tab S3 tablets with 4-GB RAM and a 9.68-inch screen with 2048 × 1536 resolution on the Android operating system version 9 to test our app. We used Unity (Unity Technologies) to develop the app and used Google’s speech service to enable speech recognition and text to speech in both English and Swedish. Moreover, we downloaded an English-language package for Google’s speech service to make the speech recognition system work with English commands even when the device is offline.

User Evaluations
Participants
We recruited some participants from the study by Shahid et al [14]. They voluntarily joined our user evaluation. Furthermore, we approached more older adults in Skellefteå, Sweden, with similar profiles as those in the iVO project. We tried to recruit people in three different roles: (1) older adult, (2) formal caregiver, and (3) informal caregiver.

Experimental Procedure
Overview
For our study, we designed the user evaluation test to run for 1 hour for each participant. This involved 30 minutes of firsthand experience using our developed app and its different UIs followed by an interview for 20 minutes. During the user evaluation, the participants freely navigated each UI, and a researcher assisted them in experiencing every feature of our app. Finally, the participants were handed a questionnaire to fill in on their own, which took approximately 10 minutes. During the evaluation, the participants interacted with the app keeping in mind their personal context of being a formal or informal caregiver to an older adult or being an older adult using such an app for themselves.

User Evaluation
Owing to the pandemic, we were limited in meeting participants from many nursing home and caregiving domains. As a result of the social distance policy, we met participants with up to 4 people at once. When we arranged a meeting with an older adult, we always grouped them with their informal caregivers or friends to make the older adult feel comfortable during the evaluation. Before starting the evaluation, we informed each participant about the process and obtained their consent. The participants were free to withdraw if they felt uncomfortable. In the user evaluation, we explained our app and the features of the UIs while the participant had firsthand experience with them. We introduced each UI in the following order to emphasize the difference between them: (1) tile view, (2) map view, and (3) AR view.

Individual Interview
After the participant had finished experiencing all the app features, we conducted an individual interview. During the interview, the conversation between participants and researchers was recorded under agreement for data analysis later. A number of questions were designed by referring to the technology acceptance model for the interview [17]. We asked about their impression and perception of the UIs and app features throughout the interview (eg, which UI was preferred based on the purpose of app use, which interaction modality helped use the preferred UI, and how easy to use and useful were those UIs and interaction modalities). We chose certain questions according to the conversation during the interviews with participants to allow for flexibility. Multimedia Appendix 4 provides a full list of interview questions.

Questionnaire
We used the User Experience Questionnaire (UEQ) designed by Laugwitz et al [32] to evaluate overall impression of the app in terms of usability and user experience. According to Laugwitz et al [32], the usability aspect comprises “efficiency,” “perspicuity,” and “dependability,” whereas the user experience aspect includes “novelty” and “stimulation.” The original UEQ contains another scale named “attractiveness” measuring another aspect of impression of the app using 6 items (ie, “annoying/enjoyable,” “bad/good,” “unlikable/pleasing,” “unpleasant/pleasant,” “unattractive/attractive,” and “unfriendly/friendly”). We omitted the attractiveness scale in our questionnaire as we were only interested in usability and user experience. As a result, we included only 5 scales (ie, “efficiency,” “perspicuity,” “dependability,” “stimulation,” and “novelty”) with 20 items in the questionnaire. The efficiency, perspicuity, and dependability scales represented pragmatic quality aspects (ie, task-related) related to usability. In contrast, the stimulation and novelty scales comprised hedonic quality aspects (ie, non–task-related) related to user experience. In the questionnaire, general information was asked about a person’s gender and age in a range. Then, 20 items were given to be answered with a 7-stage scale. Each item contained 2 opposite words, and a participant had to select a stage representing the closest scale between 2 words. The order of the words was randomized, and the order of positive and negative words was also shuffled for each item to make the participant focus on reading each item instead of selecting words with a consistent pattern. Groups of items in the same scale had similar meanings to ensure consistency in a participant’s answer. In other words, a participant’s answer could be unreliable when inconsistency arose.
Ethics Approval

This study was based on the iVO project conducted by Shahid et al [14]. The participants consented to the collection and recording of their questionnaire answers and interview data during the user evaluations. The project was, overall, in compliance with the European Union General Data Protection Regulation guidelines [33]. The data collection and processing in this study were approved by the Regional Ethical Board in Umea, Sweden (diary 2018-189/31).

Data Analysis

We conducted a statistical analysis of the questionnaire answers to identify the potential end users’ overall impression of our app. To evaluate the user experience of the 3 UIs from the questionnaire answers, we used an analysis tool provided by the UEQ team [34,35]. The analysis tool calculates means, SDs, and CIs per item and scale. The margin of error at a 95% CI was calculated by using the t value because of the sample size (N<30). In addition, a comparison of the results with those of other studies evaluated using the UEQ is presented as a benchmark. The interview answers were coded [36] to identify common impressions of participants on the 3 UIs and interaction modalities. We used inductive coding to organize data generated from observations of participants and interviews.

Results

Overview

As participants in this study were from the study by Shahid et al [14], they all had experience using a health monitoring system. In the end, we had 26 participants—17 (65%) female and 9 (35%) male. We met 96% (25/26) of participants in person, whereas we met 4% (1/26) on the web because of the limited contact owing to his job specialty during the pandemic. We removed 4% (1/26) of participants (P20) from the quantitative data because of the inconsistency in her questionnaire answers. The UEQ was used to measure the overall impression of our app, and the interviews were conducted to obtain qualitative data that could be used to understand user acceptance of the UIs and interaction modalities based on the user’s role. We categorized participants into three groups based on their role instead of their age: (1) older adult, (2) formal caregiver, and (3) informal caregiver; of the 26 participants, there were 12 (46%) older adults, 1 (4%) formal caregiver, and 13 (50%) informal caregivers. Apart from the participant whose job was as a formal caregiver, 2 (8%) participants from medical services, a nurse (P17) and a physician (P23), attended the evaluation. Most participants in the informal caregiver group (6/13, 46%) were aged from 50 to 59 years, whereas most participants in the older adult group (6/12, 50%) were aged from 60 to 69 years. All older adults (12/12, 100%) were aged >60 years, and the formal caregiver was in his 20s. Table 1 shows the information of each participant, and Figure 5 shows the demographics of the participants. The collected questionnaire data were normally distributed, as we could verify from quantile-quantile plots of the means of each scale per participant (Multimedia Appendix 5).
Table 1. Information about the 26 participants in 3 roles: older adult, formal caregiver, and informal caregiver.

<table>
<thead>
<tr>
<th>ID</th>
<th>Gender</th>
<th>Age (years)</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>Woman</td>
<td>60-69</td>
<td>Informal caregiver</td>
</tr>
<tr>
<td>P2</td>
<td>Woman</td>
<td>80-89</td>
<td>Older adult</td>
</tr>
<tr>
<td>P3</td>
<td>Woman</td>
<td>60-69</td>
<td>Informal caregiver</td>
</tr>
<tr>
<td>P4</td>
<td>Woman</td>
<td>50-59</td>
<td>Informal caregiver</td>
</tr>
<tr>
<td>P5</td>
<td>Man</td>
<td>19-29</td>
<td>Formal caregiver</td>
</tr>
<tr>
<td>P6</td>
<td>Man</td>
<td>50-59</td>
<td>Informal caregiver</td>
</tr>
<tr>
<td>P7</td>
<td>Woman</td>
<td>30-39</td>
<td>Informal caregiver</td>
</tr>
<tr>
<td>P8</td>
<td>Woman</td>
<td>70-79</td>
<td>Informal caregiver</td>
</tr>
<tr>
<td>P9</td>
<td>Woman</td>
<td>50-59</td>
<td>Informal caregiver</td>
</tr>
<tr>
<td>P10</td>
<td>Man</td>
<td>70-79</td>
<td>Older adult</td>
</tr>
<tr>
<td>P11</td>
<td>Woman</td>
<td>80-89</td>
<td>Older adult</td>
</tr>
<tr>
<td>P12</td>
<td>Woman</td>
<td>70-79</td>
<td>Older adult</td>
</tr>
<tr>
<td>P13</td>
<td>Woman</td>
<td>60-69</td>
<td>Older adult</td>
</tr>
<tr>
<td>P14</td>
<td>Woman</td>
<td>60-69</td>
<td>Older adult</td>
</tr>
<tr>
<td>P15</td>
<td>Man</td>
<td>60-69</td>
<td>Older adult</td>
</tr>
<tr>
<td>P16</td>
<td>Man</td>
<td>70-79</td>
<td>Older adult</td>
</tr>
<tr>
<td>P17</td>
<td>Woman</td>
<td>50-59</td>
<td>Informal caregiver</td>
</tr>
<tr>
<td>P18</td>
<td>Woman</td>
<td>50-59</td>
<td>Informal caregiver</td>
</tr>
<tr>
<td>P19</td>
<td>Man</td>
<td>40-49</td>
<td>Informal caregiver</td>
</tr>
<tr>
<td>P20</td>
<td>Woman</td>
<td>70-79</td>
<td>Older adult</td>
</tr>
<tr>
<td>P21</td>
<td>Woman</td>
<td>40-49</td>
<td>Informal caregiver</td>
</tr>
<tr>
<td>P22</td>
<td>Woman</td>
<td>60-69</td>
<td>Older adult</td>
</tr>
<tr>
<td>P23</td>
<td>Man</td>
<td>60-69</td>
<td>Informal caregiver</td>
</tr>
<tr>
<td>P24</td>
<td>Man</td>
<td>60-69</td>
<td>Older adult</td>
</tr>
<tr>
<td>P25</td>
<td>Woman</td>
<td>60-69</td>
<td>Older adult</td>
</tr>
<tr>
<td>P26</td>
<td>Man</td>
<td>50-59</td>
<td>Informal caregiver</td>
</tr>
</tbody>
</table>

Figure 5. The participant population by age group.
Overall Impression

Overview
We quantified 20 quality aspects that consisted of 2 words for each item in the questionnaire. We analyzed and benchmarked the responses based on the UEQ scales [34,35]. We also listed overall impressions of our app identified from the interviews.

UEQ Results

Quantified Quality Aspects
We calculated the mean, SD, and CI of each item in the UEQ that was transformed from the 7-stage scale into −3 to 3 values to evaluate the quantified quality aspects of both user experience and usability of our app. The results of each item are provided in Multimedia Appendix 6. Each item’s SD and CI were calculated from the mean of participants’ answers to each item. A total of 95% (19/20) of the items were answered over a mean of 1.6 (SD 1.19), whereas 5% (1/20) of the items (ie, “unpredictable/predictable”) were answered with a mean of 0.60 (SD 1.26). However, as the SDs for 2 items (ie, “cluttered/organized” and “confusing/clear”) were similar to their means, the differences between means and SDs were relatively smaller than for other items. Hence, we have difficulty simply accepting the results of these items as positive. In particular, “unpredictable/predictable” showed the lowest mean among all items that entered the neutral evaluation area. On the basis of CIs, some items’ results were acceptable as a positive evaluation even though they had a high SD. For example, the CI ranges for “cluttered/organized” (ie, 95% CI 0.88-2.32) and “confusing/clear” (95% CI 1.09-2.59) were >0.8, which is the minimum value for a positive evaluation, whereas those items’ means were >0.8 as well.

Scale
The mean with CI error bars for each scale is shown in Figure 6. Unlike the CIs in Multimedia Appendix 6, the CIs of each scale in Figure 6 were calculated from each participant’s mean for each scale. All scales showed a positive evaluation with a mean >1.74, and stimulation was the most valued scale. This provides evidence of positive evaluations regarding usability in terms of efficiency, perspicuity, dependability, stimulation, and novelty. When each scale was grouped into the quality aspect, the pragmatic quality aspect had a mean of 1.82, and the hedonic quality aspect had a mean of 2.17. These results represent that overall user experience in terms of task- (ie, pragmatic) and non–task (ie, hedonic)-related quality aspects received positive evaluations.

Figure 6. The mean and CI of each scale are depicted with black dots connected with lines on top of the benchmarks of each scale’s mean value, calculated from 21,175 persons in 468 studies published until 2021. The CI of each scale is calculated from the mean of each participant for each scale.

Benchmark
The UEQ team have summarized the results of UEQ from a total of 468 other researchers’ studies who also used the UEQ analysis tool. This benchmark was established from 21,175 persons’ data and is illustrated in Figure 6, along with the UEQ results for our app. We found that 2 scales (ie, “efficiency” and “perspicuity”) were rated as the second-best quality (ie, “Good”) and 3 scales (ie, “dependability,” “stimulation,” and “novelty”) were rated as the top quality (ie, “Excellent”).

Interviews

Overview
We categorized the participants’ interview data into 2 parts, and each category consisted of the following keywords. The first category contained participants’ feedback on overall impression caused by informative data, intuitive UI design, ease of use, age dependency, and lack of design clarity. The second category included user acceptance regarding the UIs and interaction modalities based on a level of ease of use and usefulness, which is presented in a separate User Acceptance section. Some interview answers that were notable for understanding participants’ perceptions of our app, UIs, and interaction modalities are provided in Multimedia Appendix 7.
Informative Data
Participants experienced that the data were informative to understand a person’s state. For example, an informal caregiver (P8) showed interest in the map view because of the supportive information for monitoring an older adult. Participants also experienced that the data on the UIs were supportive of care in a case where an informal caregiver had a problem obtaining necessary information while meeting her parent. A similar opinion was expressed by one of the informal caregivers:

Even a small event like visiting a toilet can be checked that my parent may not remember anymore. [P1]

Furthermore, another informal caregiver (P18) imagined how valuable the data could be to overcome the time and distance issues that prevented her from knowing her parent’s condition. Older adults evaluated the data as positive because of the beneficial outcomes for formal and informal caregivers. For example, an older adult (P13) thought about how useful the data could be in a specific scenario, such as when an older adult has cognitive impairment:

I can feel safer if I have this. Someone knows that I am still moving around. For instance, my children can see that I am moving. If you develop dementia, perhaps, you don’t know if you’ve eaten or not. This can tell if you did it. [P13]

One of the benefits of obtaining data for informal caregivers is that it helps understand the older adults’ states before visiting their residences (P19). The formal caregiver (P5) found that obtaining data through the tile view was preferable for him in terms of data acquisition speed and high readability.

Intuitive UI Design
Regarding the UI design, participants experienced the intuitiveness of the UIs for acquiring data. Several graphical elements were identified as helpful visual cues to aid participants in understanding the data. In the map view, the icons on buttons and the cylinders in each 3D room were perceived positively. In addition, the data visualization on the 3D map helped understand the data with spatial cues. We explained to participants that the 3D map would be the map of their residences. The data were presented in the corresponding room in the 3D map. As a result, participants experienced that the data presentation based on data-related room positions leveraged intuitiveness. For example, participants stated the reason for choosing the map view as it being a better UI than others (P10 and P11).

In the tile view, the color theme was positively received because of the improved visibility and readability of the data. For example, an older adult (P14) liked the color theme as she could obtain data by skimming through the color on each tile. When she saw the red icon on a tile, she could become aware of which activity had an abnormal behavior history before reading detailed information written in text. Different colors used for each purpose aided her in understanding the data in a short time. In addition to the design elements, the formal caregiver (P5) noted the simplicity and intuitiveness of the tile view’s layout. He found the tile view to increase the usability of the app for a caregiving service owing to quick and easy data access.

Ease of Use
Some of the participants (5/25, 20%) admitted that time was needed to get used to our app; however, 68% (17/25) of the participants explicitly mentioned how easy it was to use our app. We found these participants from all age groups and in every role. The individual preferences for UIs are unique to each participant; however, they all experienced the easiness of data acquisition.

Age Dependency
Participants felt that, even though our app was easy to use, their parents would require more time to get used to using it because of their unfamiliarity with a smartphone and app. One of the informal caregivers (P7) pointed out the different levels of user acceptance between the younger and older generation by adding an extra element, that is, a “skill,” which can be called “familiarity,” established by previous experience:

[This app is] suitable depending on the user...Not only the age but also the skills that the user has affected the experience. The younger generation can enthusiastically use it. [P7]

As evidence, we observed in the user evaluation that a relatively young adult could learn how to use AR much faster and explain it to their parent, who took more time to be able to use it by themselves. In addition, the formal caregiver (P5) showed a pessimistic perspective on the user acceptance of especially AR by older adults for the same reason that others expressed: unfamiliarity.

Lack of Design Clarity
Despite the positive experience that the app provides, some participants (6/25, 24%) experienced inconvenience from UIs caused by (1) the ambiguity of data visualization in the map view, (2) the vague motivation for use, and (3) the lack of consideration for user experience in UI design.

User Acceptance
UI Acceptance
Overview
We analyzed the participants’ UI preferences grouped by the user’s role: (1) the caregiver (ie, formal and informal) and (2) the older adult. Figure 7 illustrates the user preference for interaction modalities in the map and tile views depending on the user’s role. Personal preference could be owing to various reasons; hence, we focused on the reasons for choosing a specific UI in terms of ease of use and usefulness. Table 2 summarizes the reasons for UI preference by participant role.
Table 2. Summary of reasons for user interface (UI) preference by each participant role: older adult (O), formal caregiver (F), and informal caregiver (I).

<table>
<thead>
<tr>
<th>UI and role</th>
<th>Reason for preference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Map</strong></td>
<td></td>
</tr>
</tbody>
</table>
| O           | • Data presentation in a correlated room in the 3D map  
              • Visual representation of an older adult’s movements with trajectory lines  
              • Support for different levels of depth for data presentation |
| F           | • Data presentation tool in a meeting with others because of visual graphic components |
| I           | • More intuitive than other UIs because of the visual graphic components  
              • Overview of daily activity instead of detailed data |
| **Tile**    |                       |
| O           | • Simple UI design for easily understanding the overview of data |
| F           | • Simple UI design with informative data without unnecessary information |
| I           | • Simple UI design for fast data acquisition  
              • Familiar UI design  
              • More detailed data than in the map view |
| **AR**      |                       |
| O           | • Communication with others for social interaction  
              • As a condition report in an emergency to provide additional data |
| F           | • Active participation in health care rather than being observed |
| I           | • Additional data collection, such as facial expressions  
              • Making older adults participate in health care  
              • Feeling relieved through communicating with others |

a Impression rather than a reason for preference.
b AR: augmented reality.

**Map View**
The map view was the second most preferred UI among participants, chosen by 40% (10/25). The intuitiveness of the map was the reason that participants in all roles selected it as their preferred UI. In total, 58% (7/12) of the older adults preferred the map view for mainly three reasons: (1) the data were placed in related rooms, (2) the movement lines were...
visually presented, and (3) the data were available in both overview (see the map view in Figure 3) and detailed (see the map view in Figure 4) views. The formal caregiver proposed an idea to use the map view as a way to inform and communicate details about patients (i.e., older adults) to clients (i.e., informal caregivers). In total, 23% (3/13) of the informal caregivers wanted to see an overview of daily activity. Visual elements such as icons and cylinders on the 3D map helped them understand an older adult’s state in a short time.

Tile View

The tile view was the most preferred UI, chosen by 60% (15/25) of the participants. The principal reason for preferring the tile view was the intuitive UI design. Regarding the UI design’s intuitiveness, not only a simple UI design but also a familiar UI design could be perceived as an intuitive interface [37]. Participants in every role perceived the tile view as an intuitive and effective UI for overviewing data because of the simple design. Older adults wanted an overview of the data, and the formal caregiver preferred an overview with less detail, which was unnecessary for him. In addition, informal caregivers expressed that the tile view provided more detailed data than the map view, and they got used to the tile view because of the similar design to the app they had used before.

AR View

None of the participants selected the AR view as their preferred UI; however, most participants (10/14, 71% of caregivers and 10/11, 91% of older adults) showed interest in using the AR view as a supplementary tool for additional data mining and social interaction. Informal caregivers claimed that their parents’ facial expressions gave additional information not written in the text. Furthermore, other informal caregivers perceived that using the AR view could make them feel relieved by communicating via a facial image and activity data. Meanwhile, the older adults had a positive impression of using the AR view to communicate with their children. Sharing the captured facial image and conversing about it with others would amuse older adults who might be lonely. In contrast, sharing the captured image was perceived as a visual report for older adults to update their families on their condition. The formal caregiver declined to use the AR view; however, he saw potential use by older adults within a health care service, as did an informal caregiver (P9), because of the active participation of the older adults in their health care rather than being passively observed by others.

Interaction Modality

Overview

Our app supports multiple interaction modalities. This section analyzes the participants’ preferences for the input and output modalities. The summary of reasons for interaction modality preference is shown in Table 3. Similar to the reasons for the UI preferences, personal preferences on interaction modality could vary; hence, we focused on the reasons in terms of ease of use and usefulness. We speculate on several participants when they did not explicitly express the reason for modality preference. As the vision as an input modality in the AR view and the vibration as an output modality in every view were not principal modalities for delivering information, we excluded them from Table 3.
Table 3. Summary of reasons for input and output interaction modality preference by participant role: older adult (O), formal caregiver (F), and informal caregiver (I).

<table>
<thead>
<tr>
<th>Interaction modality and role</th>
<th>Reason for preference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Input</strong></td>
<td></td>
</tr>
<tr>
<td>Touch O</td>
<td>• Familiar modality</td>
</tr>
<tr>
<td>F&lt;sup&gt;a&lt;/sup&gt;</td>
<td>• Touch is faster than voice command</td>
</tr>
<tr>
<td>I</td>
<td>• Comfort with touching for navigating the UI&lt;sup&gt;b&lt;/sup&gt; because of many buttons</td>
</tr>
<tr>
<td></td>
<td>• Simple and familiar modality</td>
</tr>
<tr>
<td>Voice O&lt;sup&gt;a&lt;/sup&gt;</td>
<td>• Comfort with giving voice commands for navigating the UI because of ambiguous button designs</td>
</tr>
<tr>
<td>I&lt;sup&gt;c&lt;/sup&gt;</td>
<td>• Alternative modality for those who need another channel for interaction</td>
</tr>
<tr>
<td>Touch and voice O</td>
<td>• Switchable modality depends on a user’s state</td>
</tr>
<tr>
<td>I</td>
<td>• Find suitable modalities by using each of them</td>
</tr>
<tr>
<td><strong>Output</strong></td>
<td></td>
</tr>
<tr>
<td>Visual O</td>
<td>• Location-based intuitive data presentation</td>
</tr>
<tr>
<td></td>
<td>• Meaning of colors helps understand data</td>
</tr>
<tr>
<td>F</td>
<td>• Reading is faster than listening</td>
</tr>
<tr>
<td>I</td>
<td>• Quick understanding of data</td>
</tr>
<tr>
<td></td>
<td>• Familiar to read information</td>
</tr>
<tr>
<td>Audio O&lt;sup&gt;a&lt;/sup&gt;</td>
<td>• Comfortable with listening rather than reading data on the screen</td>
</tr>
<tr>
<td>F&lt;sup&gt;c&lt;/sup&gt;</td>
<td>• Alternative modality for people who want to listen</td>
</tr>
<tr>
<td>I&lt;sup&gt;c&lt;/sup&gt;</td>
<td>• Different information from that of the written text can be delivered</td>
</tr>
<tr>
<td>Visual and audio O</td>
<td>• Selectable modality depending on a user’s state</td>
</tr>
</tbody>
</table>

<sup>a</sup>Speculation based on a participant’s feedback and observations.
<sup>b</sup>UI: user interface.
<sup>c</sup>Impression rather than a reason for preference.

**Input Modality**

Finger touches and voice commands were used as input modalities. In addition, we identified some participants who preferred to use both modalities.

Touch input was preferred by 92% (23/25) of the participants, including participants who chose multiple modalities. The principal reason was that participants felt that the touch interaction was simple and familiar on a smartphone. Older adults and informal caregivers remarked on the simplicity and familiarity of touch interaction. Another reason given by one of the informal caregivers (P7) was related to the characteristic of the preferred UI. As the map view has various objects to click on for navigating data, P7 felt comfortable touching them instead of using the voice command that required memorizing every command for proper use. The touch interaction required fewer steps than the voice interaction to obtain the desired data. The formal caregiver emphasized how vital the data acquisition speed was for him. Therefore, we speculate that the formal caregiver preferred touch because of the speed of interaction.

Regarding the voice command, one of the older adults (P10) perceived the map view as better than the tile view for obtaining information. However, unlike P7, he felt that he could better
control the map view with voice command in comparison with the touch interaction. Therefore, he preferred the voice command over the touch interaction. In addition, some informal caregivers who preferred touch interaction found the value in voice command as an alternative modality for people who have an obstacle to using touch interaction. P19 noted the following:

*If you are blind, I can imagine you have a different perspective [on the value of voice interaction] than I do.*

Some participants (6/25, 24%) wished to have both input modalities for mainly 2 reasons. First, a physical impairment caused by aging or an injury changes the modality preference. In total, 33% (4/12) of the older adults, who chose both modalities, admitted that touch interaction would be the primary interaction modality when they started using the app because of its simplicity and familiarity. Meanwhile, 15% (2/13) of the informal caregivers initially wanted both modalities as they needed time to decide on the main modalities. Once they chose specific modalities as their primary interaction, they would like to stick with them.

**Output Modality**

Participants could obtain data through visual elements (e.g., icons, text, and 3D objects) and audio. Similar to the input modality, we found that some participants wanted to have both output modalities.

Visual elements as output modality were perceived positively for mainly 3 reasons. The first was intuitive data presentation with locational information. Specifically, the map view used various visual elements such as icons, lines, and cylinders on the 3D map to provide information about a person’s behavioral activities, such as transitions between rooms, overall time spent in a room, and activity in each room. Older adults experienced the strength of the visual elements on the map as they could understand information by simply seeing them instead of reading text. Another older adult (P14) reported the role of colors in recognizing data on the tile view. As long as the visual elements have a simple and understandable design, the data can be successfully delivered to participants in a relatively short time compared with audio output. The speed of data acquisition was the second reason for the preference. The formal caregiver (P5) preferred the tile view rather than the map view, mainly as reading text was fast and convenient for him. Some informal caregivers (2/13, 15%) also liked to see the data on either the map view or the tile view as they could obtain information quickly by seeing visually represented data. The last reason for this preference was familiarity with reading. People are used to reading content; therefore, many informal caregivers chose visual elements as their preferred output modality.

The audio output is the system’s feature to read text when a user presses a button. The system reads either displayed data on the screen or a summary of the data the user is seeing. An older adult (P12) preferred this audio output while using touch interaction. According to P12, even touch interaction was challenging for her; however, it was relatively more manageable than the voice command. Hence, she chose touch interaction as the main input modality. From this, we speculate that her choice of audio as a preferred output modality was made because of the relatively simple process to obtain information compared with reading. An informal caregiver (P6) pointed out that using the audio output had little merit as there was no difference in information between the written text and audio output. In other words, the participants may be willing to use audio output if there is a difference in information between the 2 different outputs. In addition, the formal caregiver (P5) found a potential use case of audio as an alternative output modality for specific users who have a reading disorder or do not want to read.

In total, 25% (3/12) of the older adults answered that they preferred having both output modalities to consider when switching between them. For example, a change in the user’s physical condition caused by aging may trigger the modality change. In other words, they considered using the 2 output modalities separately rather than simultaneously.

**Discussion**

**Principal Findings**

We used the UEQ to evaluate the initial overall impression of our app in terms of usability and user experience. The questionnaire answers regarding pragmatic (i.e., usability) and hedonic (i.e., user experience) quality aspects showed that most items from all scales were rated positively. The participants positively evaluated all the items on every scale except the dependability scale. We identified an item (i.e., “unpredictable/predictable”) from the questionnaire data analysis with a relatively low mean and high SD compared with other items in the dependability scale. This result implies that there is room for improvement regarding the unpredictable behavior of our app against the user’s expectations. However, there could be another reason for this that needs to be further investigated. This could be the participants’ different understanding of the questionnaire items [35]. This confusion could be caused by the participants’ context while taking the questionnaire. For example, the item “unpredictable/predictable” asked whether our app had reacted as the participants expected. However, several participants (5/25, 20%) asked about the meaning of “unpredictable/predictable.” In addition, we found that some participants who selected negative or neutral words for “unpredictable/predictable” chose positive words for other items on the same scale. Therefore, we assume that this result could be caused by either a misinterpretation of an item or an outlier.

The benchmark was used as complementary data to show the quality of our app, and we found that our app was rated as at least “Good” on all scales. However, the mean of the “dependability” scale was relatively lower than that of other scales. As the item (i.e., “unpredictable/predictable”) in the “dependability” scale could affect the result, we presume that an evaluation with a clear explanation and additional participants could provide more reliable results. Overall, the participants expressed interest in our app because of its usefulness for checking an older adult’s condition through intuitive UIs and ease of use with a steep learning curve [38]. The questionnaire results support the interview answers. For example, positively rated words such as “supportive,” “valuable,” “motivating,” “easy,” “understandable,” “easy to learn,” and “clear” support
the participants’ answers regarding “informative data,” “intuitive UI design,” and “easy to use.”

Although 68% (17/25) of the participants expressed that our app was easy to use, some participants (6/25, 24%) still expressed uncertainty about the UI design in terms of “ambiguous data visualization,” “vague motivation,” and “lack of consideration for user experience in UI design.” In addition, the inconvenience invoked by unfamiliarity was a noticeable phenomenon among older adults. Informal caregivers were concerned about this problem for their inexperienced parents when a new technology was introduced, such as the AR view. To resolve the uncertainty, each UI should be finely designed (1) to provide a clear meaning in visual elements, (2) to stimulate end users with a reasonable and sufficient motivation for feature use, and (3) by giving enough consideration to user experience. In addition, the learning process should be supported with media, such as video demonstrations [39], to help older adults get used to the app and UIs.

Throughout the analysis of user preference for the map and tile view, we identified “intuitiveness” and “simplicity,” the importance of which was verified by other studies [20,40,41], as the factors affecting user acceptance to a greater extent. A total of 64% (7/11) of the older adults preferred the map view as it was intuitive because of various visual elements combined with locational data, whereas the tile view impressed 79% (11/14) of the caregivers with its simple UI design. We then identified that the most preferred input modality by participants in every role was touch interaction as it was simple, fast, and familiar. Even though several older adults and a few caregivers (8/25, 32%) were interested in using the voice command, it was perceived as a secondary rather than a primary modality. Regarding an output modality, all the caregivers (14/14, 100%) liked to see the information because of fast data acquisition. Several older adults (4/11, 36%) wanted to listen, whereas 64% (7/11) still preferred to read the data from visual components.

The results of user preference for the UIs and interaction modalities go against our hypotheses. The identified reason for selecting the tile and map view were as we expected; however, both types of caregivers selected the tile view, whereas older adults were interested in the map view. We hypothesized that formal caregivers would like to use the map view for comprehensive data provided by intuitive visual components. According to the interviews, data acquisition speed was the primary factor for using our app by the formal caregiver. Therefore, obtaining information from the app should be swift and concise. Although the map view could provide fruitful data intuitively, a simple UI for fast reading of information was prioritized. As expected, informal caregivers preferred the tile view. However, as the voice command and audio output had several drawbacks, such as a necessitating learning and being slower than reading visual outputs, informal caregivers highly relied on touch interaction for as input modality and visual elements as output modality. We can mitigate the drawbacks by updating the app to understand natural languages for voice commands. Regarding the older adults’ preferences, we confirmed that touch interaction was the primary modality owing to familiarity. Unlike the caregivers, a relatively higher number of older adults (5/11, 45%) were interested in using the voice command. Output modality preference was also different from that of caregivers in that several older adults (4/11, 36%) wanted to listen because of feeling comfortable with it. Our hypothesis about older adults was incorrect as many older adults showed interest in using the voice command and audio output. Although both output modalities were perceived as secondary, having an available alternative is important because of the possibility that older adults’ state requires another modality for interaction.

The benefit of using multiple modalities is the flexibility of the interaction so that users can decide upon their preferences and states. We expect that the flexibility would enable users to have a better user experience than with a unimodal interaction modality. However, supporting multiple interaction modalities without an apparent purpose is less beneficial than unimodal interaction [27,28]. Similar to the voice command, we identified that the audio and vibration for output modality needed a redesign. As the audio output read aloud almost identical information to that on the screen, participants received the same information again, which was less valuable. To resolve this issue, we can make the audio and written information on the screen different. Essential information should remain the same; however, a slight change in the audio output could be applied for engagement. We also received several comments regarding the vibration. First, it was barely sensible because of the subtle intensity. Second, the icons on the screen already provided information that the vibration tried to notify. As a solution, we can renovate the vibration to enhance the notification with an SMS text message and push alarm. Giving a user the option to configure the amplification and repetition of the vibration can be another improvement.

During the interviews, 100% (1/1) of the formal caregivers and 31% (4/13) of the informal caregivers doubted that their parents would use the AR view. They were concerned about their parents’ low acceptance of the AR view because of unfamiliarity, health-related constraints, and complex procedures. Indeed, AR is not a familiar technology for older adults who are not even familiar with a smartphone. As the formal and informal caregivers predicted, none of the older adults chose the AR view as their preferred UI. However, 10 older adults (n=4, 40% aged 70 years and n=6, 60% aged between 60 and 69 years) perceived the AR view as acceptable to use. Overall, 84% (21/25) of the participants perceived the AR view positively, which was contrary to several caregivers’ assumptions. On the basis of the positive user acceptance of the AR view by participants aged >60 years, we presume that relatively younger generations will be more open-minded about using AR when they get older as they are familiarized with AR apps that are widely popularized, such as Facebook, Instagram, and Snapchat. In fact, we identified that 77% (10/13) of the informal caregivers perceived the AR view as useful, and the principal reason for showing interest in using the AR view among them was the informative aspect of the facial image. Informal caregivers expressed that checking their parents’ faces and reading the activity information helped them seek clues about symptoms that showed in their facial expressions. They also commented that seeing their parents’ behavior data while talking to them would be more convenient than them sending an image. Accordingly, we anticipate positive feedback on
enabling the AR view during a video call, which needs further study.

Limitations
The number of participants per user role in our study was 14 caregivers (n=13, 93% informal and n=1, 7% formal) and 11 older adults. However, the participants in each user group were homogeneous in terms of having experience with health care services. Therefore, the results of the qualitative interview analysis in each user group, especially the older adult and informal caregiver groups, were saturated with an acceptable level according to criteria from other studies [42,43]. Furthermore, the results of the overall impression of our app were reliable as an initial end-user evaluation as we recruited >20 participants of various age groups, of different genders, and in diverse roles [44,45]. In general, during the COVID-19 pandemic, we had difficulty recruiting participants for the user evaluation. In a future study, we will recruit more participants to improve the reliability of the results. In the context of measuring credibility, our participants had 1 hour to experience and evaluate our app’s design. This time constraint may have hindered the participants from having enough time to try every feature of our app in a real use-case scenario. In addition, as we aimed to evaluate initial impressions, we conducted the user evaluation without a task for measuring task-related performance. In the future, a long-term evaluation can be conducted to collect data in real life to identify the issues that influence user acceptance. This evaluation will enable us to measure the perceived usefulness through practical tasks in real life.

Our app shares the personal data of an older adult with caregivers; thus, data privacy concerns are inevitable. A participant raised an important point about a potential violation of personal privacy. Such potential conflicts regarding privacy and security can be mitigated by allowing the end user to decide what data can be shared and establishing different security layers to prevent unauthorized users from accessing the data.

Comparison With Prior Work
Health monitoring systems have been widely studied as various IoT sensors enable a system to read the contextual information of an older adult [46,47]. The objective of a monitoring system is to understand the states of persons, environments, and products based on the collected information; thereby, a service that is useful for an older adult and formal and informal caregivers could be delivered. It could be a service to aid an older adult’s daily life by providing information [48,49] or services to detect abnormal events in an older adult’s activities to inform a formal or informal caregiver [50,51]. As health care services require a number of technologies to run, user acceptance of the technologies used should be evaluated to validate their effectiveness. We found a few studies that conducted user acceptance testing on health care services; however, the target user groups were young people [52,53] rather than older adults. Moreover, other studies aimed at older adults used a stationary device at a nonindividual residence [54] or used 2D visual components only [55].

To perceive an older adult’s state precisely, it is favorable to use as many data types as possible instead of a single data type because of the different levels of richness of the identifiable information. For example, a passive infrared motion sensor could identify a person’s presence in a place; however, information from biosignals that are useful to understand a person’s physical and mental states could not be identified [56]. Pinto et al [57] even used several types of data, such as an accelerometer, room temperature, and body temperature, to track a person’s states; however, the necessity of additional sensors for collecting a person’s vital signals to monitor in-depth body conditions was stated as future work. In fact, researchers have attempted to use different types of IoT sensors to gather various types of data to understand a person’s state in detail with reliable accuracy [58,59]. The activities of daily living [60,61] and a person’s physical state [62,63] are examples of data that health monitoring systems use. Furthermore, with the growing scale of data quantity and the increasing data complexity, the data analysis method is shifting to use machine learning for improving system performance and handling large-scale data effectively [64-68]. The advantages of data diversity and machine learning adaptation in a smart health monitoring system are decent. Moreover, we found similar advantages of using multiple data with machine learning in the study by Shahid et al [14]. In the study by Shahid et al [14], various indoor sensors such as door, motion, and power plug sensors were used to collect data, whereas specific sensor data were used in the algorithm that was designed to track an older adult’s daily activities with reliable accuracy. When abnormal behavior was detected, an SMS text message was sent to a resident’s formal or informal caregiver to inform of the abnormality.

However, we wanted to go a step further than SMS text message notification for delivering information to users, including older adults and their formal and informal caregivers. As there was an explicit need for an app communicated by participants throughout the iVO project, SMS text message notification was used only for notifying abnormal events; however, our app can highlight or visualize different aspects of the older adult’s activities in detail. Well-designed data visualization could help a user understand information more quickly and easily, thereby expanding the data accessibility to those who might have an obstacle to using such a mobile health care app. Accordingly, we developed the UIs, the map view, tile view, and AR view, for our app based on the data from the study by Shahid et al [14].

Regarding the user experience, several researchers have evaluated AR on different devices, such as smartphones [9,69], tablets [26], a projector [10], and head-mounted displays (HMDs) [11,12,70], to find a beneficial aspect of using it in health care. Using an HMD sounds promising for AR apps as a camera on an HMD is always available. In contrast, other devices require extra effort, such as holding a smartphone to view and installing a camera that links to a projector for motion capture. However, we chose tablets as a target device for running our app for the following reasons. First, HMDs are uncommon in a house where an older adult lives alone. Second, as the target user group is an older generation aged >50 years, HMDs are...
inconvenient to use frequently because of the weight on the head compared with a tablet.

AR brings with it 2 strengths for use in the domain of smart health care. First, AR helps engage and motivate users to use a system continuously. Once the users become familiar with using AR, they will accept the technology. Although several researchers have conducted user evaluations of AR for older adults in health care domain [12,26] and games [71,72], there is limited research on the user experience aspect of AR with older adults in the smart health care domain. Second, AR enhances the intuitiveness of data presentation [15]. As the target user group of our study was aged >50 years, the data readability on a screen is important from a user perspective. The purpose of a health-monitoring app is to convey information correctly in an easy-to-understand manner; hence, low readability would cause inconvenience for using the app. On the basis of these strengths of AR, the face filter could be helpful for older adults in a health care scenario. The face filter is a well-known technique that combines AR and facial feature detection to overlay AR contents onto the user’s face on social networking services such as Facebook, Instagram, and Snapchat. Javornik et al [73] found that using a face filter for communication can boost social interaction between people. In addition, sending older adults’ faces to their formal or informal caregivers is equal to sending complementary data to others, thereby making older adults more actively participate in their health care [74]. Despite the verified beneficial aspects of face filters, Javornik et al [73] drew their results from a younger generation aged between 19 and 35 years; hence, user evaluations of face filters with older adults are missing. Therefore, we chose to conduct a user evaluation and examine the level of acceptance of the face filter by older adults. For displaying the sensor data, Hadj Sassi and Chaari Fourati [15] had to prepare a printed marker to display AR. However, in our work, we overcome this limitation by using AR as a face filter.

Conclusions

The need for caring services grows year by year while the resources to support them are limited. To lighten the burden on caregivers, we designed an assistive app for older adult well-being. The app supports all 3 important roles: older adults, formal caregivers, and informal caregivers. We conducted user evaluations regarding an overall impression of the app and user acceptance in terms of ease of use and usefulness of the UIs. We designed the app’s UIs using commercial apps and feedback from the participants in the iVO project. Each UI was designed to deliver data intuitively, thus enabling the user to obtain information quickly and easily. In addition, the AR is applied as a face filter to present information in a more engaging format for older adults and caregivers. Our app received a positive overall response from the user evaluation, and we identified that specific user groups preferred each UI and modality for several reasons. Accordingly, we conclude that supporting multiple UIs and interaction modalities is essential. We expect that our results will provide insight to researchers and developers on how to design an app and UI to provide a better user experience in the older adult health care domain. As future work, we intend to conduct long-term user evaluations of our app to build on end-user perspectives with specific task-based analysis.

Acknowledgments

The authors would like to thank the pilot households and their informal caregivers and the formal caregiver team from Municipality Home-care Services of the Skellefteå municipality. This work was supported by the Swedish Governmental Agency for Innovation Systems (grant 2017-02807) in a project about Internet of Things in health and care.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Flowchart of each user interface.
[DOCX File, 172 KB - humanfactors_v10i1e42145_app1.docx]

Multimedia Appendix 2
User flow of each user interface.
[DOCX File, 1883 KB - humanfactors_v10i1e42145_app2.docx]

Multimedia Appendix 3
Shared and unique features of the user interfaces.
[DOCX File, 25 KB - humanfactors_v10i1e42145_app3.docx]

Multimedia Appendix 4
Interview questions.
[DOCX File, 19 KB - humanfactors_v10i1e42145_app4.docx]
Multimedia Appendix 5
Q-Q plots of the mean of each scale per person.

Multimedia Appendix 6
Means, SDs, and CIs of the User Experience Questionnaire items.

Multimedia Appendix 7
Summary of interview answers.

References


Abbreviations

AR: augmented reality
HMD: head-mounted display
IoT: Internet of Things
iVO: Internet of Things within health and care
UEQ: User Experience Questionnaire
UI: user interface

©Joo Chan Kim, Saguna Saguna, Christer Åhlund. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 08.03.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Exploring User Visions for Modeling mHealth Apps Toward Supporting Patient-Parent-Clinician Collaboration and Shared Decision-making When Treating Adolescent Knee Pain in General Practice: Workshop Study

Simon Kristoffer Johansen, BA, MA; Anne Marie Kanstrup, BA, MA, PhD; Kian Haseli, BSc; Visti Hildebrandt Stenno, BSc; Janus Laust Thomsen, BSc, MSc, MD, PhD; Michael Skovdal Ratheff, BSc, MSc, PhD, Dr med

Center for General Practice (CAM-AAU), Department of Clinical Medicine, Aalborg University, Aalborg East, Denmark
2The Rectorate at Aalborg University, Aalborg University, Aalborg, Denmark
3Department of Health Science and Technology, Faculty of Medicine, Aalborg University, Aalborg, Denmark

Corresponding Author:
Simon Kristoffer Johansen, BA, MA
Center for General Practice (CAM-AAU), Department of Clinical Medicine
Aalborg University
Fyrkildevej 7, 2nd floor
Aalborg East, 9220
Denmark
Phone: 45 60788895
Email: skjohansen@dcm.aau.dk

Abstract

Background: Long-standing knee pain is one of the most common reasons for adolescents (aged 10-19 years) to consult general practice. Generally, 1 in 2 adolescents will continue to experience pain after 2 years, but exercises and self-management education can improve the prognosis. However, adherence to exercises and self-management education interventions remains poor. Mobile health (mHealth) apps have the potential for supporting adolescents’ self-management, enhancing treatment adherence, and fostering patient-centered approaches. However, it remains unclear how mHealth apps should be designed to act as tools for supporting individual and collaborative management of adolescents’ knee pain in a general practice setting.

Objective: The aim of the study was to extract design principles for designing mHealth core features, which were both sufficiently robust to support adolescents’ everyday management of their knee pain and sufficiently flexible to act as enablers for enhancing patient-parent collaboration and shared decision-making.

Methods: Overall, 3 future workshops were conducted with young adults with chronic knee pain since adolescence, parents, and general practitioners (GPs). Each workshop followed similar procedures, using case vignettes and design cards to stimulate discussions, shared construction of knowledge and elicit visions for mHealth designs. Young adults and parents were recruited via social media posts targeting individuals in Northern Jutland. GPs were recruited via email and cold calling. Data were transcribed and analyzed thematically using NVivo (QSR International) coding software. Extracted themes were synthesized in a matrix to map tensions in the collaborative space and inform a conceptual model for designing mHealth core-features to support individual and collaborative management of adolescents’ knee pain.

Results: Overall, 38% (9/24) young adults with chronic knee pain since adolescence, 25% (6/24) parents, and 38% (9/24) GPs participated in the workshops. Data analysis revealed how adolescents, parents, and clinicians took on different roles within the collaborative space, with different tasks, challenges, and information needs. In total, 5 themes were identified: adolescents as explorers of pain and social rules; parents as supporters, advocates and enforcers of boundaries; and GPs as guides, gatekeepers, and navigators or systemic constraints described participants’ roles; collaborative barriers and tensions referred to the contextual elements; and visions for an mHealth app identified beneficial core features. The synthesis informed a conceptual model, outlining 3 principles for consolidating mHealth core features to support individual and collaborative management of knee pain.

Conclusions: An mHealth app for treating adolescents with knee pain should be designed to accommodate multiple users, enable them to shift between individual management decision-making, take charge, and engage in role negotiation to inform...
shared decision-making. We identified 3 silver-bullet principles for consolidating mHealth core features as enablers for negotiation by supporting patient-GP collaboration, supporting transitions, and cultivating the parent-GP alliance.

(JMIR Hum Factors 2023;10:e44462) doi:10.2196/44462

KEYWORDS
mobile health; mHealth; design; patient physician relationship; collaborative care; shared decision-making; adolescents; parents; knee pain; patellofemoral pain; Osgood Schlatter; musculoskeletal; general practice; primary care; mobile phone

Introduction

Background

Approximately one-third of adolescents (aged 10-19 years) experience long-standing musculoskeletal pain [1,2]. At the age of 10 years, there is a sharp increase in adolescents consulting their general practitioner (GP) owing to a musculoskeletal complaint [3]. Between the age of 12 and 15 years, musculoskeletal complaints are the fourth most common cause for consulting general practice [4]. The most common pain site is the knee, accounting for between 30% and 50% of all presentations in this age group [5]. Previously, knee pain was considered to be benign and self-limiting [6], but recent studies have demonstrated that 40% to 50% of adolescents still struggle with pain after 2 to 5 years [7], and 7 out of 10 had reduced or halted their sports participation owing to knee pain [8]. This is problematic because life course studies highlight adolescence as a transition period [9], where health habits are formed and carried into adulthood [10]. Interventions with exercises and leaflets with patient education on managing knee pain have shown potential for improving adolescent’s prognosis [11-14], but maintaining adolescents’ performance with exercise and self-management activities remains as a barrier to success in this patient group [11]. Ensuring that adolescents learn to effectively self-manage their knee pain is important to enable patients to gain corrective experiences while reducing the period with experienced limitation [15,16].

Mobile Health Apps

Mobile health (mHealth) apps are promising tools for improving the treatment of adolescents with everyday management needs owing to chronic conditions [17,18], and their acceptability of the technology is considered to be high owing to their common use of mobile phones [19]. Defined as “health practices supported by mobile and wireless devices” [20], mHealth apps draw upon the always-present, always-on properties of smartphones [21] to deliver just-in-time health interventions, text reminders, tailored information, self-tracking, connectivity, and decision-making support to contexts where patients experience their conditions [22]. Although literature highlights how the inclusion of mHealth may foster more patient-centered treatments [23], the evidence for mHealth’s efficacy in creating positive health outcomes and behavior change in patients remains indicative and contradictory [24]. Systematic reviews (2013-2020) have documented how including mHealth apps in the treatment of adolescents were associated with improvements in disease awareness, self-management abilities, treatment adherence, psychological well-being, and behavior change across conditions (asthma, diabetes, arthritis, and psychological issues), but the findings were inconsistent [25-33]. Qualitative studies support that mHealth can assist the development of personal management strategies and assist young patients in engaging with clinicians in cocare situations [19,34,35]. However, very few provide guidance on how future mHealth apps should be modeled to be integrated into complex treatment settings as tools for enhancing existing treatments and facilitating continual care [36-38].

Self-management and Shared Decision-making

Self-management is essential for achieving recovery from knee pain [39]. Clinician-delivered patient education has been hailed as effective for teaching patients to self-manage their knee pain [13,40], but health literacy studies point to adolescents’ here-and-now perspective on injuries, their desire for independence, and capacity for understanding GPs’ instructions as barriers when supporting young patients [41,42]. Involved parents can help adolescents’ transition to self-management, through task assistance, coaching, guidance, rewards, and help during management mistakes [43-46], but this requires agreement on tasks and responsibilities to enable collaboration [43-47]. Shared decision-making holds the power to engage multiple stakeholders in the planning and facilitation of care, by merging patient and caregiver preferences with evidence-based practices [48], and the concept is central to the collaborative care process [49]. Exploring the visions of adolescents with long-standing knee pain, parents, and GPs about how an mHealth app can support individual and collaborative self-management when adolescents receive GP treatment for their knee pain may identify targets for designing mHealth tools, which are easy to be integrated into existing treatment practices, and empower identification and resolution of management and adherence barriers through shared decision-making [37].

Objectives

This study aimed to identify principles for designing mHealth core features, which are sufficiently robust to support adolescents’ everyday management of their knee pain and sufficiently flexible to act as enablers for supporting patient-parent-GP collaboration and shared health decision-making.

Methods

Study Design

Action research was included as a methodological framework to guide our application of methods, analysis, and knowledge production [50,51]. The project’s intervention component consisted of 3 future workshops [52,53]: 1 with young adults with knee pain since adolescence, 1 with parents of adolescents

https://humanfactors.jmir.org/2023/1/e44462

JMR HUMAN FACTORS | vol. 10 | e44462 | p.112

(page number not for citation purposes)
with knee pain, and 1 with general practice physicians. Participants’ dialogues were captured via audio recorders and analyzed separately using reflective thematic text analysis [54] to map the general challenges and visions of each participant group for an mHealth app. The extracted insights and visions were synthesized in a matrix, to identify lanes of collaboration and tension sources and to facilitate the crystallization of design principles [55]. The study was reported in accordance with the guidelines in the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist, to ensure that our communication of findings corresponded to domain specific standards (Multimedia Appendix 1 [1-4,6-9,13,19,34-61]).

Ethical Considerations
The study protocol was submitted for revisions to the regional board of research ethics in Northern Jutland, and they ruled that the project was permitted to continue without registration based on national guidelines.

Participants
We included 3 study populations that were separated in terms of their roles in the clinical setting—patients, clinicians, or next of kin. Young adults (aged 18-25 years) with long-standing recurring knee pain during adolescence (emerging age 10-15 years; duration >6 months) were included into study population 1. The decision to include young adults was rooted in how self-management skills are developed over time [56] and how young adults would be capable to critically reflect on how mHealth features could have supported their transition to self-management. Exclusion criteria included competing musculoskeletal or pain conditions, severe physical handicaps, psychological issues, and surgery of the knee. Finally, GPs were included into study population 3. Inclusion criteria were employment in general practice for at least 1 year, experience in treating adolescent knee pain, and willingness to participate. Participants for study populations 1 and 2 were recruited through social media posts targeting individuals in Northern Jutland, containing the link to a form with questions related to the inclusion criteria, contact information, and consent forms. Potential participants, who expressed interest in participating and consented to contact, were contacted via phone by SKJ; screened; provided informed about the project, participants’ rights, and data treatment procedures (oral and written); and recruited. Participants in study population 3 were identified within the Center for General Practice and Nord-KAP—the Quality Unit for General Practice in Northern Jutland’s clinician networks, contacted via email and phone, informed, screened, and recruited into the project.

Future Workshops
We drew upon the future workshop as described by Jungk and Mullerts [62] as a template for planning our study’s intervention component. A key feature of the future workshop relates to the use of coconstruction of knowledge, through collaborative activities and ideation to extract novel and useful solutions to complex real-world problems. Future workshops use a 3-step process, entailing critique, fantasy, and defining shared visions [52] to guide participants’ dialogues toward formulating shared visions for possible futures (refer to Table 1).

To facilitate this transition between the workshop phases, a generative activity was designed, which used case vignettes [63] and inspiration cards [64] to encourage participants’ dialogues and guide them through to the third and final phase of our future workshops. Furthermore, it was decided to forgo the final phase (follow-up phase), because implementation was outside our scope of inquiry.

### Table 1. Overview of the future workshop phases

<table>
<thead>
<tr>
<th>Timeline and phases</th>
<th>Brief explanation about the phases</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Before</strong></td>
<td></td>
</tr>
<tr>
<td>Preparation</td>
<td>Organizers and facilitators agree on the theme, invited participants, methods, location, locales, rules, and timetables of the future workshop.</td>
</tr>
<tr>
<td><strong>During</strong></td>
<td></td>
</tr>
<tr>
<td>Critique</td>
<td>Participants investigate the problem through criticism and brainstorming. Challenges and ideas are noted and organized into themes.</td>
</tr>
<tr>
<td>Fantasy</td>
<td>Participants create a picture of a utopic future. Brainstorming and creative techniques are included to suspend criticism and extract ideas.</td>
</tr>
<tr>
<td>Implementation</td>
<td>Participants organize, evaluate, and develop ideas related to practicality and ease of implementation. Action plans are developed.</td>
</tr>
<tr>
<td><strong>After</strong></td>
<td></td>
</tr>
<tr>
<td>Follow-up</td>
<td>Action plans are monitored, changes are performed, and new future workshops are planned to address challenges to implementation.</td>
</tr>
</tbody>
</table>

*aPhases and descriptions are adapted from a paper by Vidal et al [53].

bThe follow-up phase is not included in this paper.
Case Vignettes
A case vignette (Multimedia Appendix 2) was designed in collaboration with young adults with knee pain, parents, and GPs to outline salient features of young patients, seeking treatment for their knee pain in general practice. The case was included to initiate discussions, by presenting a patient narrative with relevant and irrelevant information [63]. The case was tested and iterated with 3 GPs, 2 parents of adolescents with knee pain, and 2 young adults with knee pain to ensure comprehension.

Inspiration Cards
An inspiration card exercise was developed to encourage dialogue and cocreation and guide participants through the future workshops’ 3 phases (Multimedia Appendices 3 and 4). The inspiration card game featured themes related to 3 conceptual domains: physical domains, problems or challenges, and possible solutions. The themes were identified by 3 members of the project group (SKJ, MSR, and JLT) and reflected our initial understanding of the challenges and experiences encountered by members of all 3 study populations. The inspiration cards were tested and iterated following the same pattern as the case vignettes to ensure relevance of themes and comprehensibility.

Other Artifacts
Participants in each group were provided with other artifacts such as post-it notes in 3 colors (red, yellow, and green), pens, and pen markers, which participants could use to brainstorm ideas; organize emerging themes to visualize conceptual relations; and engage in a shared evaluation of themes, concepts, and novel ideas. To support participants in bridging the gap between ideation and visioning (future workshop phases 2 and 3), the facilitators advised groups to rearrange, explore, and prioritize their ideas by reorganizing the design cards before engaging in the work of phase 3.

Setting and Procedure
Special care was taken to ensure that all 3 workshops followed the same procedure to heighten the compatibility of the extracted insights and visions. Workshops 1 and 2 (young adults and parents) were conducted at a local community center, whereas workshop 3 was conducted at the Center for General Practice in Aalborg. All workshops lasted approximately 3 hours, distributed across three 40-minute phases and brakes. Each workshop was conducted with a primary coordinator (SKJ), a workshop facilitator who introduced workshop activities and guided participants through the 3 phases, and 2 cofacilitators (AMK, MSR, and JLT) who would help the facilitator in guiding group discussions and otherwise observe the process from the background. Upon arrival, participants were divided into work groups of 3 to 4 participants each. Each workshop was initiated with a short introduction by the facilitator and a presentation by an invited specialist, physiotherapist, mHealth specialist, and eHealth specialist. The facilitator would then introduce the case vignettes and the inspiration cards corresponding to the given phase and provide instructions about how to complete the exercises of each phase. This procedure was repeated before each of the 3 phases (critique, ideation, and vision phases) of the workshops. Each phase was concluded with a plenary discussion, during which the groups presented their thoughts and ideas, while the facilitator summarized key points on a flipboard and asked follow-up questions. Upon completion of the final phase, all groups presented their visions for an mHealth app for feedback from other participants and facilitators. The workshops concluded with a debriefing session, during which the participants were informed about their rights, completed the consent forms, and were given the opportunity to ask final questions.

Data Collection
Overall, 3 types of data were collected to illuminate the problem from different angles. Clinical characteristics of study populations 1 and 2 were collected using web-based forms, whereas core data from population 3 were collected through phone interviews. During workshops, participants’ visions and insights emerging from plenary discussions were noted on flipboards by the facilitator, and group discussions were captured via audio recorders for analysis and interpretation using reflexive thematic analysis (RTA).

Analysis
The data gathered during the 3 future workshops were analyzed through RTA by Braun and Clarke [54] by the lead researcher (SKJ) and 2 student workers (KH and VHS). The data sets from each individual workshop were transcribed for meaning retention, as described by Kvale and Brinkmann [65], using Expresscribe (NHC Software) transcription software. The transcribed data sets were analyzed in parallel through a 4-stage process including familiarization, coding and identification of themes, condensation and refinement, and synthesis into a shared narrative (Multimedia Appendix 5). NVivo (version 11; QSR International) coding software was used for the coding and organization of themes, and coding lists were created and maintained by all coders (SKJ, KH, and VHS) during each individual analysis. Identified themes were refined through iterative cycles of horizontal readings and condensation, and related subthemes were merged. Emerging thematic overlaps, divergencies, and relationships identified within the individual analysis were discussed among the lead researcher (SKJ) and student workers (KH and VHS) until consensus was reached. Considering the thematic relationships identified during each individual analysis (Multimedia Appendix 6), a narrative of 5 storyboard themes was identified. Consecutively, the insights uncovered during each thematic analysis were organized in a matrix to map collaborative tensions, identify design principles, and inform a conceptual model [55]. To ensure coding integrity, stakeholder checks were conducted by the lead researcher (SKJ) and student workers (KH and VHS), and coding list entries were discussed as the analysis progressed. The lead researcher (SKJ) was responsible for the final abstraction and presentation of the findings in the 5 storyboard themes, which outlined the narrative in the Results section. All involved parties (SKJ, KH, and VHS) approved the storyboard themes, narrative, matrix analysis, and conceptual model before the analysis was concluded.
Results

Inclusion of Participants

The social media posts and phone screening generated 36 potential participants for workshop 1 (young adults) and 19 potential participants for workshop 2 (parents). Our efforts to contact GPs in the Northern Jutland area via emails and cold calling yielded 17 potential participants from the 21 who were initially contacted (Figure 1). From the 11 included young adults, 9 (82%) participated (n=8, 89% women; mean age 20, SD 1.73; range 18-23 years), whereas 2 (18%) withdrew their participation. All participants in workshop 1 (9/9, 100%) had experienced long-standing knee pain, emerging between the age of 11 to 16 (mean 13, SD 1.32) years and lasting for an average of 5.8 (SD 2.45; range 3-9) years. From the 11 parents included, 6 (55%) participated (n=5, 83% women; mean age 46, SD 1.44; range 41-52 years). Parents reported how all their adolescents (mean age 11, SD 1; range 10-12 years) had experienced knee pain for an average of 2 (range 1-6) years and how 67% (4/6) of participants had consulted their GP, 33% (2/6) had consulted physiotherapists, and 33% (2/6) of the parents had not yet sought treatment for their child’s knee pain. From the 12 included GPs, 9 (75%) participated (n=4, 44% women; mean age 42, SD 11.84; range 30-63 years), whereas 3 (25%) canceled in advance. The participants in workshop 3 had an average of 8.5 (SD 7.8; range 1.5-25) years of experience in general practice, with 58% (7/12) of them reporting having a special interest in musculoskeletal conditions.

Figure 1. A flowchart providing an overview of the 3 lines of inclusion from when participants responded to our outreach efforts (social media posts or emails).

Results of the Data Analysis

Overview of Themes

RTA uncovered a narrative of 5 storybook themes. Overall 3 themes described the roles participants played within the treatment situation, and 1 theme described the collaborative barriers and challenges across contextual settings. Theme 5 identified core features and collaborations based on the participants’ visions for an mHealth app. The insights from the analysis were summarized within a matrix to inform a conceptual model, identify principles for expanding the design of mHealth core features, and enable patient-parent-GP collaboration and shared decision-making.

Theme 1—Adolescents as Explorers of Pain and Social Rules

The first theme comprised statements describing how participants experienced their emerging knee pain and the challenges related to the everyday management of knee pain. The analysis revealed how young adults described their emerging knee pain as fluctuating or something that emerged in different situations such as stair climbing, bicycling, running, sports, and gym class and affected the adolescent’s ability to engage in valued activities. The young adults described being tasked with exploring ways to cope with emerging pain and pain-related frustration, while managing the social consequences of being limited. A common theme during workshop 1 was how emerging knee pain initiated a vicious cycle in which adolescents’ efforts toward minimizing the social consequences...
or hiding their condition resulted in increased pain. A young adult described the cycle as follows:

_There is this vicious circle, where you start feeling the pain during sports, talk about it at home, go to the doctor, and then the doctor tells you that there is nothing. You go back to everyday life again, return to school, try to spend time with your friends and start to feel pain again. You withdraw for a little while, and start being left out of your [friend] group. So, you return to sports to get back into the group and the cycle continues._ [Participant 3]

One of the main challenges described, related to the invisibility of knee pain and how the adolescents were dependent on others recognizing their pain (peers, parents, and coaches) to avoid being branded as lazy, whiney, or careless. Another recurring theme related to how fear of being “benched” may result in adolescents “forgetting” or ignoring their pain to fulfill social obligations or avoid exclusion. A participant described how she considered ignoring her pain to avoid missing out of activities with friends:

_Well, it is possible that you might forget to tell others about your [knee] pain, because you’re afraid that they’ll think you’re a cry-baby or that you won’t be allowed to participate in things you’re normally allowed to...We were in Africa one winter, and the team went to climb a mountain and I had to wait by the foot. Back then I considered not telling [the others] that I had pain, so I could come along._ [Participant 5]

Alternatively, I group (workshop 1) described how acceptance from others or honesty about the knee pain was important and empowered adolescents to stop hiding the pain and focus on managing the condition. A participant articulated the link between gaining parents and GP’s acceptance and managing the knee pain in other situations:

_It’s more like a step on the journey towards gaining this acceptance from the world, but when you have the backings of your parents and the doctor...I think that makes it easier to manage [knee pain], I definitely remember, how it was easier to manage [the knee pain], when my mother was involved._ [Participant 2]

Apart from acceptance, participants described performing regular knee exercises and learning to “find the limit” with their knee pain as essential for breaking the vicious cycle and balancing self-care while performing everyday activities and how this was challenging for adolescents. Although young adults and parents highlighted adolescents forgetting their knee pain in nonimpact situations or losing faith in exercises as barriers to breaking the negative spiral, young adults and GPs emphasized how learning to differentiate between good and bad pain is essential for managing the knee pain:

_I was always told that I just needed to be warmed up, so I ended up thinking doing sports was equal to having knee pain, and therefore I never really learned to find the limit where I should have stopped in relation to the pain I felt. The result...I would come home from training and have to lie down with my leg up because I was in pain._ [Participant 9]

Finally, young adults, GPs, and parents highlighted how adolescents may struggle to remember and expressing their pain in words when asked by parents or GPs. This posed a challenge when reaching out to parents or health care professionals for support in managing their knee pain. One group of participants (workshop 1) highlighted fear of stigma as a contributor to this problem, whereas another group’s (workshop 1) comments indicated that adolescents lacked the vocabulary for describing their pain beyond the immediate pain experience. A participant (workshop 1) described it in following way:

_I also found it difficult when my physio would ask the question; where do you get pain, what is it that cause you pain, and what does it feel like?...I don’t know, because in this moment I don’t have any pain. So I can’t give you an explanation on how the pain is..._ [Participant 5]

### Theme 2—Parents as Supporters, Advocates, and Enforcers of Boundaries

A recurring theme during the second workshop was parents referring to taking on the parent role to solve a problem. Although the parents generally recognized the adolescents as individuals with their own opinions, experiences, and desires, the parent role term was often used in recognition of how certain aspects of management were difficult for adolescents. Thus, parents had to step in and take control to remove barriers that inhibited adolescents’ management of their knee pain. A parent described the parent role in terms of supporting their child:

_My role as a parent is to take her [daughter] seriously...To do the right thing a hundred percent...this includes seeking out everything [treatments] to find out exactly what this [the knee pain] is. To back her up 100% percent, all the way through the health system._ [Participant 14]

Another participant exemplified how taking the parent role also meant stepping in and setting boundaries when they felt their adolescents were not able to do so themselves. A parent described setting limits for her son’s participation in soccer:

_Sure, I could tell my son that [he had to take brakes], but I’m sure he would just lie his way out of it. Well, he can’t do that right now, because I often accompany him during training and matches...You’re not match ready, since I am the one deciding this...Still, it’s hard to keep them away from it [sports] because this is what they are really keen on doing._ [Participant 10]

Although parents from both groups highlighted being present, listening, taking complaints seriously, and setting boundaries as important for supporting their adolescent’s health decisions, both groups described alternating among the 3 tasks of emphasizing with the adolescents, advocating for the adolescents, or reassessing their own understanding of knee pain to support their adolescents by creating situations where adolescents were capable of self-management. Furthermore, parents described having experienced how their adolescents

---

Johansen et al

JMIR Hum Factors 2023 | vol. 10 | e44462 | p.116

(page number not for citation purposes)
struggled to remember, understand, and express their knee pain in words. This complicated the parents’ task of assessing when to seek treatment, resulting in parents overlooking or negating adolescents’ attempts to express their knee pain. Thus, the parents had to learn to read between the lines, within the adolescents’ descriptions. This need was exemplified in a parent’s description of discussing the knee pain with her daughter:

I always had to ask my daughter how bad is it? She doesn’t really complain about it [knee pain] except for what she tells me when we were in these situations...and then she’ll just tell me: but mom, I just think I’ve gotten used to it [knee pain]. [Participant 11]

In terms of advocacy, parents in both groups described instances where they had stepped in and negotiated on their adolescents’ behalf and how this advocacy initially occurred in the clinical setting and extended into the parents and adolescents’ networks after consultations. Parents described that negotiation with the GP aimed at supporting adolescents in articulating their pain and ensuring that their child benefited from their consultation. Parents described how advocacy also included withdrawing from treatment or seeking alternative treatments and information sources if they felt invalidated or that the GP did not meet their needs during consultations. A parent described how her expectations had prevented her from seeking additional treatments for her daughter:

During spring we had a longer period where I thought that we might have to take her to a GP [for the knee pain], but where our own GP who we have been seeing for years ended up quitting...and I just thought why bother because then would have to see a new one. I know the old GP would have taken it into account if I told him that we had waited and seen for a long time. We had waited for three months...But it was not him anymore so I thought we wouldn’t bother. [Participant 10]

**Theme 3—GPs as Guides, Gatekeepers, and Navigators of Systemic Constraints**

The data from workshop 3 revealed several tasks, responsibilities, and dilemmas, which GPs had to navigate when treating adolescent knee pain. GPs described taking on the role as teachers or coaches, tasked with guiding the adolescents into a positive spiral with decreased somatization; better disease management; and sustained, balanced participation in sports as a positive spiral with decreased somatization; better disease management; and ensuring that their child benefited from their consultation. Parents described how advocacy also included withdrawing from treatment or seeking alternative treatments and information sources if they felt invalidated or that the GP did not meet their needs during consultations. A parent described how her expectations had prevented her from seeking additional treatments for her daughter:

During spring we had a longer period where I thought that we might have to take her to a GP [for the knee pain], but where our own GP who we have been seeing for years ended up quitting...and I just thought why bother because then would have to see a new one. I know the old GP would have taken it into account if I told him that we had waited and seen for a long time. We had waited for three months...But it was not him anymore so I thought we wouldn’t bother. [Participant 10]

The participants’ statements during workshop 3 indicated how treating adolescents with knee pain was a 2-step process and how ruling out serious pathologies or trauma, diagnosing the condition, identifying the right treatments, informing, and managing expectations was part of the initial step of treatment. A GP described how identifying alarm symptoms was important:

Yes, we need to know the alarm symptoms...Are there any symptoms we professionally know that; “Oh this, this we need to effectuate on immediately if we spot it.” We need some kind of screening feature for what is acute, what is dangerous and not. We are doctors, that’s why patients come to us in the first place. [Participant 16]

Besides momentary management, the analysis revealed how GPs developed and used different behavioral strategies in tandem with usual care, to encourage the adolescents to explore, gain insights, and gradually become better at making health decisions going forward. A GP described the strategies he used to supportive strategies:

Something that could be really beneficial is to explain to people how the pain emerges...I sometimes use the term “Pain memory,” that you can have pain on an injury that is almost fully healed, but you will continue to experience pain right? So sometimes it can be useful to show them that it [their knees] can’t break. Some people have a belief that things may like fall apart. [Participant 17]

Although one GP group highlighted how this required understanding the “full patient,” other strategies included addressing the patients’ worries and concerns to facilitate acceptance, encouraging trial and error by providing suggestions for managing pain fluctuations, exercising, motivating adherence via goal setting, and establishing working alliances with parents. However, GPs highlighted that their efforts toward supporting future self-management depended on whether adolescents felt that following the GP’s advice allowed them to better understand and control their pain. A GP described the importance of understanding the whole patient with the following quote:

You need to look at the “whole patient” like her [the case],...what she wants to archive. I normally differentiate between the lazy bodies and the non-lazy ones. With the lazy bodies, the problem is often that they will state that they have pain, because they might stand to gain from it...like I can’t participate in gym class or bicycle to school. [Participant 17]

GPs identified several constraints within the treatment situation that had a direct influence on the GP’s treatment decisions and possible outcomes. The long disease course of knee pain and patients slipping through the cracks were highlighted as the main concerns and challenges of the GPs. Moreover, GPs pointed to adolescents’ difficulties in articulating their knee pain and patients or parents’ misunderstanding of GP’s instructions as barriers, which contributed to adverse outcomes such as dissatisfaction, withdrawal from treatments, and parents
insisting for surgery. A GP described how knowing when his message got across to the adolescents was a challenge. Other GPs suggested that forming a therapeutic alliance with parents could help facilitate the knowledge translation, avoid withdrawal, and provide GPs with the ability to monitor and adjust their treatments by proxy. A GP elaborated further on this, by stating how maintaining adherence was ultimately the patients and parents’ responsibility:

But the problem is all too real in the clinic, as a lot of things will disappear within a short time, but that again means that we should be better to provide patients a safety net if it [knee pain] continues. But they also have a responsibility for coming back again [if pain persists]. I can’t take on that responsibility, all the time. [Participant 19]

**Theme 4—Collaborative Barriers and Tensions**

Our workshops uncovered how adolescents, parents, and clinicians engaged in different types of collaboration aimed at empowering adolescents to enter an upward spiral with increased understanding of the disease and self-management. The analysis identified several communicative barriers, which lead to tensions in parent-patient-GP communication. Young adults, parents, and GPs highlighted adolescents’ difficulties in remembering and verbalizing their knee pain as a major source of tension. GPs described how this acted as a barrier to fulfilling their role in terms of diagnosis, management education, and planning future treatments, and parents and young adults corroborated this, with young adults suggesting that adolescents’ inability to explain their knee pain may be related to lack of in-depth understanding of their knee pain. A GP suggested how pain diaries could be used to alleviate tensions, by helping adolescents to articulate developments in their pain:

Because, most 10- to 15-year-old adolescents, when you ask them to recall; How many or how often do you experience knee pain, which time of the week or whatever this might be, will have a hard time providing an ample description of this, so this way you may get an overview of how they are impacted by their knee pain...And maybe it could be combined with something [a feature] which gives an indication of their pain severity. [Participant 18]

Another tension source was related to adolescents or parents’ expectations of obtaining a solution to the knee pain when entering a treatment collaboration. This was corroborated by young adults and parents, who described how being told “wait and see” could lead patients and parents to conclude that the GP did not believe them or know how to treat their knee pain. A parent described feeling invalidated after receiving the “wait and see” recommendation for her adolescent, and this had affected her expectations of future consultations:

Now the two of us are here where we haven’t quite made it to the GP’s yet...and we have discussed it and believe it boils down to us feeling that we weren’t heard when it...and being sent home and told to “wait and see.”...so the thought of us being sent back home again...well then, we might as well wait and see [by] ourselves. [Participant 10]

GPs corroborated this during workshop 3, by highlighting how they knew there were limits to what they could do for adolescents consulting with knee pain and how managing parents’ expectations was quintessential when gatekeeping, to avoid parents becoming frustrated and seeking other treatments prematurely. GPs also discussed using imagery to give parents something tangible, build alliances, avoid withdrawal, and prevent parents from insisting for surgery:

It depends on how the parents are involved in this...If you can’t get an alliance with them before you have made a scanning and they are just like a white wall...like they’re simply not listening, and you know they’ll eventually walk out the door and seek out a private clinic or something, then I might open up the possibility of getting a scanning, but I generally believe that it [scanning] may potentially do more harm then good, because you might find something [unrelated]. [Participant 18]

Furthermore, GPs highlighted systemic constraints such as consultation times, subpar IT systems, and loss of communication owing to referrals as barriers leading to loss of contact with patients and parents:

We discussed how the condition may persist for a long time, potentially without a whole lot of doctor-patient contact...So when we are first made aware of the injury until they return...it could be months, even years apart before we see the patient again. And we haven’t had a chance to affect the outcome, apart from a few weeks’ time. [Participant 20]

Parents described how their lack of knowledge about knee pain caused tensions when assessing whether additional treatment was merited, when advocating with the GP, and when communicating their adolescents’ conditions and forming alliances with actors in their networks (teachers, coaches, and other parents). Thus, parents and young adults described how parents’ lack of knowledge meant that they risked overlooking or negating adolescents’ symptoms, unnecessarily restricting their sports participation, or accepting nonbeneficial treatments. A parent suggested how tailored patient information could help parents to know when to seek additional treatments:

I would have liked having a guide for how long it takes...I know sundhed.dk [Danish government health portal] has something where you can describe your symptoms and whatever, and in the bottom I know they have something like...now its lasted for so and so long, and then should do this and this. If it looks like this, you need to contact your GP...like a guide of some sorts. [Participant 12]

Parents and young adults described an emerging dynamic, which eventually led to the formation of an alliance in which parents helped adolescents to create space for their self-management in everyday situations. Although young adults described how parental recognition made it easy to confront teachers, peers, and coaches about knee pain, reaching out to parents meant...
risking being dismissed or restricted from sports participation, which created tensions. In contrast, ensuring their child’s well-being was highlighted as quintessential to parents, but their lack of understanding about knee pain sometimes led them to take the wrong actions when their adolescents presented pain. Parents described how taking a more trusting approach reduced tensions and allowed them to focus on gatekeeping: advocacy during GP visits; and engaging with teachers, coaches, and other parents to create space for their adolescent’s self-management. A parent articulated this in the following way:

But it comes back to what responsibility you have as a parent. Because you make the decision to enter actively into it [supporting the adolescent] and provide your input. And by this I don’t mean entering something into a dead system. You look the other person into the eyes and say: I have this issue with my child, can we work out a solution together. [Participant 14]

**Theme 5—Visions for an mHealth App**

Finally, the analysis identified several visions for mHealth core features for enhancing collaboration and shared decision-making across collaborative spaces. The visions were distributed across 3 categories, directed toward enhancing reassurance, supporting pattern recognition and articulation of knee pain, and enhancing 2-way communication. However, participants described these core features as intersecting and needed to support different activities simultaneously for maximum effect. A GP described how his group envisioned that an mHealth app should support different tasks simultaneously:

I’m thinking that you could create a three-legged system. Like something with monitoring of, what’s the status of this [knee pain]. How is it developing. A tool for treatment as well as a patient education tool. [Participant 20]

Participants envisioned an mHealth app containing features for reassuring adolescents and helping them to test and evaluate their management decisions, when the knee pain emerged in everyday situations. Both GPs and young adults suggested how a library (videos) with trustworthy information about knee pain mechanisms, possible trajectories, and a first aid kit for managing flareups could reassure adolescents, promote self-education, and allow them to share this knowledge with peers. However, the young adults suggested that patient cases with other adolescents with knee pain were more easily relatable for youths and could provide hope for betterment. Furthermore, GPs and young adults highlighted that adolescents sometimes struggled to remember and comply with exercise programs and that adding a library with exercise videos and in-depth explanations could reassure adolescents that they were performing the exercises correctly. Ensuring that exercises were actionable (short and easily understandable) while combining them with a tracking feature could motivate exercise adherence by visualizing the short-term and long-term effects of exercises:

Well, I did actually get started on some type of rehabilitation, but I eventually quit because I didn’t really feel that it worked...so if you’re thinking apps, then incorporating one [a feature] which provides you with suggestions for exercises and gives you reminders like “remember to make these exercises.” [Participant 9]

Participants across all groups suggested having core features that empowered adolescents to monitor, explore, and identify patterns in their knee pain. Both young adults and GPs described how a journal feature could support adolescents’ self-management by helping them in identifying activities that caused pain. Nevertheless, young adults and parents suggested incorporating reminders and predefined pain scales to reduce the burden related to monitoring the knee pain. All participants suggested that visualizing journal entries could help adolescents in overcoming their challenges by remembering pain-causing activities and articulating their knee pain when it emerged. Young adults suggested how incorporating a map visualizing common developments in the knee pain could assist adolescents and parents in assessing how the knee pain progressed and establishing treatment goals. This was corroborated by GPs who described how this feature could help adolescents in identifying activities that would not affect their knee pain:

...And then there was something with a pain measurement [feature], where you could note it as logboook with where you had pain and how much pain you had, but a combination of them, where you could get the connection between...I have this pain, maybe it subsides when I’m not active. [Participant 1]

Finally, participants envisioned how core features could be expanded to enable negotiation of meaning and shared decision-making, but this required a balance because actors had different information needs. Participants generally agreed that the journal and visualization features were central to this, by providing GPs and parents’ insights into the adolescent’s experience. Young adults and GPs described how visualizing journal entries could help resolve tensions in GPs and adolescents’ communications during clinical visits by providing a common ground for discussions, whereas GPs and parents described how visualizations could also help GPs in adjusting treatments to the patient’s needs. However, GPs described how this required visualizations to be aggregated for easy overview to avoid time loss. Another vision related to the exercise library was how incorporating a checklist with symptoms to look out for could help parents in deciding when to seek additional treatments and prepare parents for engaging with coaches, teachers, GPs, and physiotherapists. Finally, GPs and parents described how an mHealth app could facilitate information flow during transitions between treatments or when negotiating with external actors (physiotherapists, coaches, and teachers), by alleviating tensions related to parents or adolescents forgetting information obtained from clinicians between consultations. This was corroborated by the young adults, who exemplified how an app could facilitate an ongoing negotiation among multiple actors, to ensure acceptance of the knee pain. This was exemplified in the following quote:

We discussed how it [an app] should be a little bit like “school-parent communication software”...but in a way where you have communication between the patient and the GP, and where the GP can post your...
exercises along with comments, videos or whatever... This way we get that acceptance of how the pain is real, which means the surrounding world are in on accepting them, but you need to start with the ones who are closest, like mom and dad. [Participant 2]

Results of the Matrix Analysis

The Conceptual Model

The matrix analysis informed the construction of a conceptual model. Organizing participants’ descriptions about their roles, tasks, challenges, and interactions within a system identified how adolescents, parents, and GPs were interconnected within a triadic relationship, where all actors engaged in different modes of management behaviors (Figure 2). Considering the identified tension sources, the model outlined targets for designing mHealth core features to bridge the gap between the supporting participants’ individual management practices and collaboration across multiple contexts.

Figure 2. The conceptual model that was designed to illustrate the complex interplay between participants’ roles, their proximal and distal goals, management tasks, and barriers present in the collaborative space. GP: general practitioner.

Decision-making and Negotiation

The layout with the embedded triangles illustrated that the participants’ collaboration in managing the adolescents’ knee pain unfolded at the individual and community levels across multiple contexts. A key insight was not only how participants took on different roles, tasks, and responsibilities within the collaborative space but also how these roles were often dual-sided and contradictory in nature. The individual triangles (top, left, and right) were designed to illustrate how the actors (adolescents, parents, and GPs) navigated these role-based contradictions via their management decision-making (center) in their individual contexts—an act that was obscured to other actors unless disclosed via words or observable actions. The matrix analysis identified how all actors encountered management barriers, which they could not resolve themselves (eg, obtaining a diagnosis, gaining knowledge about knee pain, and securing social support) and caused tensions in the collaborative space (Multimedia Appendix 7). To overcome these barriers, actors engaged with other actors to draw upon...
their competencies (as adolescents with knee pain, parents, and GPs) to expand their management capabilities (decision-making), modify contexts (eg, being excused from gym class and creating a working alliance), or adjust their roles and tasks in the collaborative space. However, participating in these exchanges meant renegotiating the participants’ individual goals, roles, and tasks to be effective (center and inner triangles). When successful, the negotiation may strengthen the actor’s individual health decision-making capabilities, articulate shared goals, and cultivate working alliances. Nevertheless, failure to negotiate was identified as having a cascading effect, leading to increased tensions and complexity in the collaborative space, inhibiting shared decision-making, and prompting treatment withdrawal if the tensions were not resolved.

**mHealth Core Features as Collaborative Enablers**

Considering participants’ visions for an mHealth app, described challenges, and identified tension sources (Multimedia Appendices 7 and 8), the matrix analysis identified several touchpoints where participants interacted to resolve individual management challenges and how these interactions contributed differently toward sustaining the collaborative situation. This informed 3 principles for organizing mHealth core features as collaborative enablers for supporting tension reduction by empowering negotiation and informing shared decision-making.

**Enhancing Communication**

The young adults, parents, and GPs envisaged how health information collected via quantified self-tracking could support adolescent-GP communication and how their information needs differed in terms of timing, timelines, and modalities. Participants described how the act of self-tracking knee pain via, for example, pain journals, receiving tailored patient information (etiology and exercise support), and performance feedback, could help adolescents in assuming the role of explorers through the identification and articulation of patterns in their knee pain. However, this required that the delivered health information should be actionable in everyday settings to encourage exploration, compliance, and articulation. Furthermore, visualizations of aggregated self-tracked data could help adolescents and GPs to overcome communicative barriers by assisting adolescents in recalling and articulating previous developments in their knee pain, while simultaneously giving GPs a foundation for guiding the adolescents—by providing GPs an overview of the adolescents’ trajectory, the ability to monitor the effects of treatments and exercise regimes, and a starting point for discussing future treatments. However, effective presentation and delivery of the self-tracked health information were crucial to ensure GP use in complex clinical settings.

**Facilitating Transition**

Young adults highlighted how acceptance and adapting an honest perspective about knee pain was important for facilitating adolescents’ transition to self-management and how parental support could help adolescents to take on the role as explorers. The analysis revealed how different types of static information (patient cases, lists of symptoms, exercise videos, and patient information) could promote safety in making management decisions (individual level) by providing reassurance, along with vocabulary and expert information that adolescents could include when explaining their knee pain to peers, coaches, and GPs to avoid stigma. For parents, static information delivered with the app (eg, leaflets or webpages targeting parents and adults) could empower them to create space for adolescents’ exploration of their knee pain (decision-making) and remove management barriers through negotiations with other parents, teachers, coaches, and GPs. This included enhancing parents’ knowledge about knee pain symptoms and treatment types, while providing them with guidance and tools for how to engage and educate other actors, coaches, teachers, and other parents about knee pain and the management needs of their child.

**Forming Alliances**

Finally, the analysis outlined how communicative difficulties between GPs and parents could lead to tensions and parents deciding to withdraw and seeking other forms of treatments and how this was driven by parents not feeling seen or heard when consulting GPs. Both parents and GPs suggested incorporating core features that could help parents and GPs in entering negotiation and building alliances. Including a checklist for parents with symptoms and questions for GPs could limit tensions by ensuring that parents felt heard during consultations, while providing GPs space for addressing parental expectations to treatments. Furthermore, providing parents with information materials (folders and webpages) about the adolescent’s symptoms, treatment options, and prognosis could help them to adjust their expectations, while cultivating a sense of co-ownership and forming treatment alliances with the GP.

**Discussion**

**Principal Findings**

Our findings revealed several key insights that should be considered when designing mHealth apps as tools for facilitating patient-centered treatment of adolescents with knee pain in general practice. Our analysis indicated how adolescents, parents, and GPs entered a triadic relationship with different goals, tasks, and information needs, similar to what Hohmann [57] and Brooker [58] observed in pediatric settings. Participants worked toward 1 outcome—ensuring that the adolescent entered an upward spiral with decreasing pain and increasing control. Adolescents actively facilitated this transition at the individual level, through their exploration of their knee pain in the present [56,59]. In contrast, parents and GP’s roles were peripheral and focused on supporting adolescents in navigating future management obstacles through observation, encouragement, boundary setting, and provision of management advice and information. Our analysis showed how all actors alternated between 2 modes of management behaviors. This included making individual management decisions to overcome contextual management challenges and engaging with other actors to use their expertise (as adolescents, parents, and GPs) to adjust management practices or collaborations—an act that involved a renegotiating of goals, tasks, and responsibilities to be successful. Although negotiation acted as a linchpin for shared decision-making [60], our analysis identified how articulation, lack of knowledge, unfulfilled expectations, and...
nonreciprocity inhibited negotiation and increased tensions in the collaborative space. Participants envisioned how an mHealth app for adolescents with knee pain should focus on providing reassurance, pattern recognition, and facilitating 2-way communication. Our conceptual model identified 3 principles for expanding the design scope from supporting adolescents’ individual management decisions toward arranging mHealth core features as enablers for empowering adolescents, parents, and GPs to shift their focus from individual management toward reducing tensions via negotiation and shared decision-making, by enhancing communication, facilitating transition, and building alliances in the collaborative space.

Comparison With Previous Studies
Systematic reviews describe how including mHealth apps during the treatment of adolescents with self-management needs from chronic conditions was associated with a host of observable benefits, which included positive changes in patients’ disease understanding, self-management capabilities, treatment adherence, and health behavior [25,27,29,32,66]. However, reviews with a clinical focus emphasized how understanding the patient’s disease-specific challenges and mHealth’s position within the treatment ecology is crucial for realizing the technologies’ potential [27,29] and leveraging patient-centered care [23]. Feasibility studies have documented how the efficacy of mobile and web-based interventions with tailored information, patient education, and self-directed exercises was similar to that of face-to-face consultations when treating knee pain in youths and adults [67]. However, these studies provided little insight about how adolescents integrated the information into their everyday management practices and beliefs, which was highlighted as important by GPs and young adults.

Qualitative studies by Slater et al [35,68,69] and Stinson et al [34,70] corroborated several of our findings about how mHealth-delivered health interventions (reminders, quantified self-tracking, and data feedback) held the potential for supporting adolescents with chronic pain through awareness, acceptance, and health decision-making between consultations, which could alleviate communicative obstacles during consultations. These studies focused on using mHealth data for enhanced communication during consultations as a driver for behavior change but provided little insights into how core features should accommodate the nonlinear, context-sensitive nature of mHealth interventions [38] or how to include parents as informal carers between GP consultations [71]. Systematic reviews by Moon et al [72] and Slater et al [68] corroborated participants’ visions about how mHealth could improve communication between GP clinics and home environments to facilitate collaborative care. However, both reviews highlight how adjusting apps to be integrated into formal and informal tasks and workflows of GPs and patients is crucial to ensure meaningfulness and continual use of mHealth and other digital patient education concepts.

Adolescents’ Self-management as a Leveled Activity
Our analysis identified how adolescents’ management of their knee pain was a leveled activity, as described by Modi et al [73], which manifested within the individual and social domains. The descriptions obtained from young adults and GPs indicated a recognition of the knee pain as something processual (eg, a negative spiral), with parents and GPs being tasked with empowering adolescents to transition from a negative to a positive spiral with increased adaption, self-reflection, management, and control. These findings are similar to the observations made by Lerch and Thrane [74] in adolescents with other chronic conditions. The young adults’ descriptions indicated how adolescents’ main challenge was to balance managing their pain in situ, while simultaneously managing the social impacts of the knee pain [75]. Simultaneously, taking action to balance their knee pain allowed adolescents to explore, expand, and progress their self-management as illustrated by Johansen et al [56] and Cartwright et al [76]. This insight aligned with observations from Kralik et al [77] and Price [78] about how re-establishing balance or order in the wake of emerging symptoms acted as a point of learning for patients with chronic illness. Thus, we believe that a future mHealth app for adolescents with knee pain should target behaviors that re-establish an equilibrium with the knee pain, to enhance adolescents’ ability to facilitate their inquiry into their knee pain, which will improve their self-management in time.

Core Features for Supporting Management Decisions

Providing Actionable Advice
Qualitative studies outline how managing knee pain is complex and involves adolescents balancing several activities including understanding their pain, maintaining function in everyday situations, care seeking, self-reflecting, and adjusting to a life with pain [56,59]. Our analysis identified targets for designing mHealth apps for supporting adolescents’ efforts toward managing their knee pain between consultations. For increasing adolescents’ understanding of pain, participants suggested incorporating a first aid kit feature, with information about the etiology of knee pain and actionable advice for pain alleviation. This would provide adolescents with reassurance, while promoting reflections about pain triggers and pain relief strategies, similar to what Rathleff et al [12] documented in clinical trials with their activity management tool.

Promoting Adherence
Regarding participating in care activities, participants highlighted how remembering and sustaining correct performance with exercises was a barrier, as observed by Faber et al [79] in youths. Participants suggested that incorporating a video feature with patient information, exercise instructions, and prompts for reminders could reduce maladaptive beliefs and empower exercise adherence. This aligns with the findings of Selhorst et al [80] and Holt et al [81], who showed that video-based handouts with patient education and exercise support limit maladaptive beliefs and boost adherence in youths with musculoskeletal pain. In addition, Riel et al [82] documented how live feedback increased the compliance of adolescents with knee pain with digital exercises. Studies exploring the use of gamification elements (leader boards, goals, minigames, leveling, esthetics, feedback, and rewards) to sustain long-term adherence to health interventions have shown positive effects such as enhancing medical adherence, disease understanding, and physical activity in adults and older people with chronic conditions [83-85]. Additional user-centered studies...
are needed to ensure that gamification elements align with adolescents’ needs to exert control of their knee pain [86] and are experienced as useful, meaningful, and inspiring for self-management in everyday contexts [83].

Enhancing Articulation
Finally, recall and articulation of knee pain were highlighted as barriers when seeking support from GPs and parents. Participants suggested that goal setting, quantified self-tracking, and pain journals could help adolescents in overcoming this barrier, by registering pain triggers, identifying pain thresholds, and assessing the value of behavior change. This resonated with the findings by Slater et al [69] about how adolescents were capable of setting personal recovery goals and the descriptions by Moon et al [72] and Laloo et al [87] about quantified self-tracking’s potential for enhancing adolescent-GP communication during clinical encounters, by allowing the adolescents to capture, explore, and articulate connections among management experiences, reflections, and outcomes [88]. However, our findings expanded upon this by illuminating the gap between ensuring that the self-tracked data are actionable to the adolescents and integrate into the GP’s decision-making process to facilitate mutual articulation. In total, our insights extend the current knowledge base and inform which individual management challenges should be addressed by mHealth core features in future mHealth concepts.

mHealth Apps as Tools for Collaboration

Individual and Shared Processes
The young adults highlighted how GP and parental support had helped them to accept their knee pain and take on the role as explorers [34] and how experiencing having their requests rejected—real or perceived—increased doubts, dissatisfaction, and stagnation and prompted withdrawal, as observed in adults and youths with chronic pain [89-91]. Our analysis revealed how entering into treatment led to the emergence of a complex triadic relationship, where patients, parents, and GPs took on different roles, tasks, and responsibilities, similar to what Brooker [58] and Hohmann [57] outlined in the collaborative care triangle and what Kanstrup et al [92] presented as a complex interplay. Nonetheless, our findings expanded upon this by describing how all participants were engaged in individual reflective processes and navigated both proximal and distal goals, as illustrated by Ryan and Sawin [93] in their family self-management concept. Furthermore, the emerging triadic relationship shared several properties with the communities of practice by Lave and Wenger [94] in terms of how participants organized, collaborated, and shared knowledge and how successful participation solidified the collaborative relationship. Ensuring that adolescents entered into an upward spiral acted as a shared domain of interest [95], with our analysis indicating how this acted as a cornerstone for shared health decision-making, as described by Makoul and Clayman [96]. Still, this required that said goals were articulated to be effective.

In terms of legitimate peripheral participation, our analysis outlined how adolescents were engaged in a situated learning process, which was driven by their explorative approach to managing knee pain [95]. In contrast, parents and GPs took on the roles as masters by observing, encouraging, diagnosing (GP), informing, educating (GP), and setting boundaries to facilitate the adolescents’ transition to self-management, as observed by Cha et al [43]. However, the absence of a shared vocabulary and repertoire for addressing self-management challenges made it difficult for participants to establish a working collaborative relationship. A future mHealth app should focus on creating a shared language for addressing the processual, proximal, and distal goals of participants [93] to support shared learning and avoid early treatment withdrawal.

Negotiation and Shared Decision-making
A key insight was how participants navigated the collaborative space through 2 modes of management behaviors, echoing the descriptions of Brooker [58] about members alternating between professionalism and providing care. Although adolescents, parents, and GPs navigated individual management challenges via their management decision-making, our matrix illustrated how participants solicited the expertise of other actors to overcome individual or contextual management barriers—an act that was collaborative in nature and required renegotiation of goals, roles, and tasks to succeed. Literature highlights that negotiation is an essential component in shared health decision-making [60,96]. Our matrix analysis identified how adolescents’ articulation, memory for exercises, parent’s knowledge, the long trajectory of knee pain, and GP’s ability to engage with adolescents inhibited negotiation and shared learning and increased tensions within the collaborative space. On the basis of these insights, our conceptual model presented 3 principles for designing or consolidating mHealth core features to act as enablers for mediating between participants to support negotiation, reduce tension, and create a basis for shared decision-making across multiple contexts. The model identified targets for enhancing the interpretive flexibility of core features to accommodate multiple user needs simultaneously [97] to enhance patient-GP communication, support parents in facilitating the adolescents’ transition, and help parents and GPs in building mutual trust and enhancing patient alliances.

Relations to Existing mHealth Concepts
Our exploration of mHealth literature related to this study failed to uncover mHealth concepts that incorporated all 3 principles for supporting negotiation and shared decision-making simultaneously; 3 designs were identified, which included 1 or 2 of the previously mentioned principles. The PainApp described by Kounpourou [98] supports patient-clinician communication, by using quantified self-tracking and aggregated data to empower carers’ clinical reasoning when negotiating treatments with adults with musculoskeletal pain. Furthermore, the PainApp uses self-tracking notifications as an actionable component to encourage patients to track and reflect on treatment effects at home. The ICanCope concept by Stinson et al [70] used a hybrid design to support adolescents with musculoskeletal pain to transition to adult care. The ICanCope app uses theory-informed interventions (self-tracking, goal setting, coping skills training, and social support) that are actionable to enhance adolescents’ symptom exploration and management decision-making at home. The web component included education features with...
self-tracked data to support adolescents’ articulation and features (discussion boards, goal sharing, and self-advocacy skills) that we interpret could enhance adolescents’ safety during patient-clinician communication and when negotiating roles and space with parents. As no mHealth tools were identified with features that directly supported parent-adolescent collaboration, the diabetes management plan tool [99] was included to exemplify how mHealth core features may bridge adolescents’ needs for safety from sanctions and parents’ needs to frame and facilitate their child’s self-management when making shared management decisions. By allowing users to choose and share items representing adolescents and parents’ tasks and obligations, the tool inspires role negotiation and shared decisions by articulating the tasks and responsibilities of both parties.

Clinical and Design Implications

Literature describes how mHealth apps could act as a silver bullet for introducing patient-centered treatment approaches [23]. Our findings confirmed mHealth’s potential for augmenting collaborative care but illustrated how several bullets are required to leverage an mHealth apps’ utility as a tool for supporting shared decision-making when treating adolescent knee pain in general practice settings.

Several studies outlined mHealth’s potential for improving patient-clinician communication during consultations [35-37,68,72], but little knowledge exists on how mHealth apps could be used to build patient-GP relationships across time—a core theme in collaborative care [49]. Our analysis outlined how GPs took on the role as change agent during consultations, which entailed alternating between the expert and guide role to guide adolescents toward independence in management. However, our findings indicated that this process was dual reciprocal and required efforts from GPs to manage adolescents and parents’ expectations and biopsychosocial understanding of knee pain to alleviate collaborative tensions. Our analysis confirmed the observations by Brown et al [41] and Sawyer et al [9] about how adolescents’ articulation, ability to recall their pain developments, and memory for GP instructions acted as barriers to diagnosis, monitoring, adjusting treatments, and educating adolescents in managing their knee pain. mHealth apps with quantified self-tracked data could empower GPs to adapt measurement-based care when evaluating and adjusting interventions [36,72]. Furthermore, the act of tracking pain, reviewing aggregated data, and making management decisions could help adolescents to construct and articulate theories about how the knee pain progressed in time [88], which could then be discussed and qualified by the GP [36] to inform negotiation and shared decision-making [60]. Our analysis identified parents’ potential for taking on the role as informal carers and supporting the integration of treatments and management advice [71]; however, gaining the insight needed to know when to step in, set boundaries, and mediate between the adolescent and GPs required time, trust, and acceptance of the division of labor within the collaborative space.

The conceptual model outlined how adolescents, parents, and GPs required different modalities of information to sustain their roles, management practices, and inform negotiation. This places substantial demands on ensuring the interpretive flexibility of core features to act as enablers for shared decision-making. We infer that a future mHealth app should include 3 data loops: 1 with actionable small data interventions to support adolescents in exploring and balancing their knee pain [56]; 1 with static information for parents about etiology, red flags, and assistance for engaging with GPs; and 1 with aggregated mHealth data that allow GPs to step into the expert role and be effective in delivering treatments, adjusting treatments, and taking the coach role to provide self-management education. Ensuring that the app and information integrate into the ecologies of workflows, systems, and demands of general practice was crucial for achieving this effectiveness and should be explored further in future studies.

Strengths and Limitations

The workshops’ inclusion of generative methods for facilitating dialogue and coconstruction of knowledge enabled us to extract the tacit and latent knowledge of our participants, which may not have been accessible via qualitative or focus group interviews [100]. However, by drawing upon the lived experiences of participants, the workshops were made vulnerable to recall and saliency bias [101,102]. Furthermore, having people working in groups may have made the process more open to say-do problems, compared with single-person interviews [103]. This was addressed by incorporating plenary discussions after the workshop phases and creating a pleasant atmosphere during the workshops [104]. A key strength was how each workshop followed the same design and used similar tools (case vignettes and inspiration cards) to facilitate discussions, which created a foundation for data synthesis. To avoid thematic reproduction, the design card’s themes were kept open and participants were encouraged to expand upon them throughout the workshops [105]. Although the workshops’ production of novel insights indicated that our efforts were successful, the extracted themes and models should be viewed as symbolic, ideal representations of the participants’ shared experience and should only serve to inform scientific inquiry. Despite our efforts to balance our workshop populations, the participants were predominantly women, which may have resulted in male-specific perspectives being overlooked during workshop discussions. Furthermore, no in-depth data were collected about the participants’ socioeconomic status, making it uncertain whether all potential user demographics have been represented. Literature highlights that the ideal number of participants for 1 workshop is between 8 and 16 [61], which we were able to accommodate in workshops 1 and 3. Despite the alternating sampling sizes, all workshops produced rich and descriptive data sets, with novel insights that could inform future mHealth tools. Thus, we did not interpret the low number of participants in workshop 2 as a limitation.

Conclusions

mHealth apps are often hailed as a silver-bullet solution for introducing patient-centered and collaborative care approaches in complex care settings. Although actors navigated the complexity of the collaborative space through 2 modalities of management, role negotiation acted as a linchpin for reducing collaborative tensions and expanding actors’ management...
practices via shared decision-making. An mHealth app for treating adolescents with knee pain should accommodate multiple users and enable them to shift between individual management; take charge; and engage in negotiation of goals, roles, and tasks to inform shared decision-making. Our conceptual model identified 3 silver-bullet principles for consolidating mHealth core features as enablers for negotiation of goals, tasks, and roles by supporting patient-GP collaboration, empowering parents to facilitate transition, and cultivating the parent-GP alliance.

Acknowledgments
The study presented in this paper is independent work, which is supported and financed by the Research Foundation for General Practice (A1819; SKJ). The study was conducted at the Center for General Practice, Department of Clinical Medicine at Aalborg University. The views expressed in this paper are those of the authors and are not to be attributed to the Research Foundation for General Practice, Center for General Practice at Aalborg University or Aalborg University. The author group would like to extend their gratitude to all the young adults, parents, and general practitioners (GPs) who participated in their workshops and shared their personal stories and perspectives with them. In addition, the authors thank Nord-KAP—The Quality Unit for General Practice in Northern Jutland, for allowing them to use their network for recruiting GPs; Camilla Ulfkjær Østergaard for helping to facilitate their workshops; Andreas Lind Johansen for help with designing the conceptual model; and Negar Pourbordbari for helping the authors to advertise their study to young GPs.

Conflicts of Interest
None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 3469 KB - humanfactors_v10i1e44462_app1.pdf ]

Multimedia Appendix 2
Case vignette.
[PDF File (Adobe PDF File), 95 KB - humanfactors_v10i1e44462_app2.pdf ]

Multimedia Appendix 3
Examples of inspiration cards.
[PDF File (Adobe PDF File), 269 KB - humanfactors_v10i1e44462_app3.pdf ]

Multimedia Appendix 4
Overview of inspiration card themes.
[PDF File (Adobe PDF File), 88 KB - humanfactors_v10i1e44462_app4.pdf ]

Multimedia Appendix 5
Overview of steps taken during the data analysis.
[PDF File (Adobe PDF File), 48 KB - humanfactors_v10i1e44462_app5.pdf ]

Multimedia Appendix 6
Themes obtained from the thematic analysis.
[PDF File (Adobe PDF File), 170 KB - humanfactors_v10i1e44462_app6.pdf ]

Multimedia Appendix 7
Tension map.
[PDF File (Adobe PDF File), 249 KB - humanfactors_v10i1e44462_app7.pdf ]

Multimedia Appendix 8
Themes identified from plenary discussions.
[PDF File (Adobe PDF File), 134 KB - humanfactors_v10i1e44462_app8.pdf ]

References


Abbreviations

CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
GP: general practitioner
mHealth: mobile health
RTA: reflexive thematic analysis
Using Intervention Mapping and Behavior Change Techniques to Develop a Digital Intervention for Self-Management in Stroke: Development Study

Alex W K Wong1,2,3, PhD, DPhil; Mandy W M Fong2, PhD; Elizabeth G S Munsell1,4, PhD; Christopher L Metts5, MD; Sungmoon I Lee6, PhD; Ginger E Nicol1, MD; Olivia DePaul8, OTD; Stephanie E Tomazin1, MPH; Katherine J Kaufman1, BA; David C Mohr3,9,10, PhD

1Center for Rehabilitation Outcomes Research, Shirley Ryan AbilityLab, Chicago, IL, United States
2Department of Physical Medicine and Rehabilitation, Northwestern University Feinberg School of Medicine, Chicago, IL, United States
3Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL, United States
4Center for Education in Health Sciences, Feinberg School of Medicine, Chicago, IL, United States
5Department of Pathology and Laboratory Medicine, Medical University of South Carolina, Charleston, SC, United States
6Manning College of Information and Computer Sciences, University of Massachusetts Amherst, Amherst, MA, United States
7Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, United States
8Memorial Hospital Belleville, Barnes-Jewish/Christian HealthCare, Belleville, IL, United States
9Center for Behavioral Intervention Technologies, Northwestern University Feinberg School of Medicine, Chicago, IL, United States
10Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, United States

Corresponding Author:
Alex W K Wong, PhD, DPhil
Center for Rehabilitation Outcomes Research
Shirley Ryan AbilityLab
355 E Erie St
Chicago, IL, 60611
United States
Phone: 1 312 238 1742
Fax: 1 312 238 4572
Email: wwong@sralab.org

Abstract

Background: Digital therapeutics, such as interventions provided via smartphones or the internet, have been proposed as promising solutions to support self-management in persons with chronic conditions. However, the evidence supporting self-management interventions through technology in stroke is scarce, and the intervention development processes are often not well described, creating challenges in explaining why and how the intervention would work.

Objective: This study describes a specific use case of using intervention mapping (IM) and the taxonomy of behavior change techniques (BCTs) in designing a digital intervention to manage chronic symptoms and support daily life participation in people after stroke. IM is an implementation science framework used to bridge the gap between theories and practice to ensure that the intervention can be implemented in real-world settings. The taxonomy of BCTs consists of a set of active ingredients designed to change self-management behaviors.

Methods: We used the first 4 steps of the IM process to develop a technology-supported self-management intervention, interactive Self-Management Augmented by Rehabilitation Technologies (iSMART), adapted from a face-to-face stroke-focused psychoeducation program. Planning group members were involved in adapting the intervention. They also completed 3 implementation measures to assess the acceptability, appropriateness, and feasibility of iSMART.

Results: In step 1, we completed a needs assessment consisting of assembling a planning group to codevelop the intervention, conducting telephone surveys of people after stroke (n=125) to identify service needs, and performing a systematic review of randomized controlled trials to examine evidence of the effectiveness of digital self-management interventions to improve patient outcomes. We identified activity scheduling, symptom management, stroke prevention, access to care resources, and cognitive enhancement training as key service needs after a stroke. The review suggested that digital self-management interventions,
especially those using cognitive behavioral theory, effectively reduce depression, anxiety, and fatigue and enhance self-efficacy in neurological disorders. Step 2 identified key determinants, objectives, and strategies for self-management in iSMART, including knowledge, behavioral regulation, skills, self-efficacy, motivation, negative and positive affect, and social and environmental support. In step 3, we generated the intervention components underpinned by appropriate BCTs. In step 4, we developed iSMART with the planning group members. Especially, iSMART simplified the original psychoeducation program and added 2 new components: SMS text messaging and behavioral coaching, intending to increase the uptake by people after stroke. iSMART was found to be acceptable (mean score 4.63, SD 0.38 out of 5), appropriate (mean score 4.63, SD 0.38 out of 5), and feasible (mean score 4.58, SD 0.34 out of 5).

**Conclusions:** We describe a detailed example of using IM and the taxonomy of BCTs for designing and developing a digital intervention to support people after stroke in managing chronic symptoms and maintaining active participation in daily life.

*(JMIR Hum Factors 2023;10:e45099) doi:10.2196/45099*

**KEYWORDS**
mobile health; digital intervention; technology; SMS text messaging; intervention mapping; behavior change; self-management; stroke; rehabilitation; mobile phone

**Introduction**

**Background**

Stroke is the leading cause of long-term disability in the United States. The American Heart Association estimates that 795,000 individuals experience a new or recurrent stroke each year in the United States, and the annual cost of stroke is US $45.5 billion [1]. Globally, the prevalence of stroke is projected to affect 77 million people by 2030 owing to reducing stroke mortality rates and an ever-aging population [2]. Many people live with residual disabilities after stroke, such as reduced mobility and limitations in performing household chores and community activities [3]. Stroke often manifests uncontrollable chronic conditions (eg, diabetes and hypertension), and people after stroke also experience ongoing chronic symptoms (eg, depression and fatigue). The inability to manage chronic conditions and symptoms may lead to decreased participation in prestroke roles and activities, exacerbating other chronic diseases and increasing the chance of stroke recurrence and mortality [3,4]. Moreover, the impact of stroke on personal health and health care will continue to grow, forcing an urgent demand for innovative person-centered approaches in stroke care [5].

In stroke rehabilitation, an effective approach to addressing these long-term consequences is teaching self-management skills for people after stroke [5,6]. Self-management is defined as an individual's behavioral management of symptoms, treatment, physical and psychosocial consequences, and lifestyle changes inherent in living with a chronic disease [7]. Self-management can be delivered in various formats; the most common delivery format is via group-based face-to-face structured education programs [8]. A meta-review of 13 systematic reviews has demonstrated that self-management interventions improve daily activities, independence, and mortality in people after stroke [9]. National research agendas, such as those of the Department of Health and Human Services [10] and the Institute of Medicine [11], include self-management as a key strategic framework for promoting health, managing chronic conditions, and preventing disability.

Although self-management programs are recommended, evidence has shown that the uptake of these structured education programs is low [12,13]. Work obligations, time commitments, and low perceived benefits are commonly identified barriers contributing to the low uptake of these programs [12]. Fortunately, growing evidence supports using technologies for self-management and other behavior change interventions [14]. Digital interventions that use a broad range of technologies, such as mobile phones, the web, and sensors, have become a new means of delivering care and offering support for users in changing behaviors and cognitions related to health, mental health, and wellness [15]. A systematic review of reviews indicates that SMS text messaging interventions effectively support self-management of chronic diseases because of their high reach and highly accessible, relatively low-cost communication strategy [16]. Moreover, the American Heart Association states that digital interventions have the potential to revolutionize self-management by promoting patient and clinician engagement in active real-time care partnerships [17]. Indeed, some barriers associated with face-to-face programs, such as transportation, location, and time, are easier to address with digital interventions because users can access digital interventions at their own time and place [18].

Despite digital interventions for self-management becoming more available, the descriptions of the intervention development process are often unclear [19] and do not consistently incorporate both the patient and clinician perspectives; moreover, none exist for people after stroke. The findings of a recent systematic review regarding digital self-management interventions in low back pain have supported similar arguments. The authors found that the articles included in the review were heterogeneous and did not report intervention details, making it difficult to understand what might work best, for whom, and in what circumstances [20]. Intervention mapping (IM) provides a well-defined framework for systematically developing, implementing, and evaluating behavior change interventions [21,22]. IM emphasizes the participation of stakeholders and provides a structure for integrating theory, findings from empirical literature, and information collected from the target populations. The IM approach has been applied in designing digital self-management interventions in other chronic
conditions, such as type 2 diabetes [18] and low back pain [23], as well as nondigital self-management programs [24-26].

**Objectives**

This paper provides a use case of the step-by-step IM process used to adapt an evidence-based face-to-face self-management education program, Improving Participation After Stroke Self-Management (IPASS), which consists primarily of group-based psychoeducation [27], into an interactive Self-Management Augmented by Rehabilitation Technologies (iSMART) intervention delivered via a mobile device. We chose to adapt IPASS, which is a stroke-specific self-management intervention targeting the development and practice of chronic disease management skills to improve functioning and participation in home, work, community, and social activities. Digital tools such as iSMART can address the common barriers to participation in traditional face-to-face self-management programs, such as time, transportation, or geographic location, that limit access to, and uptake of, self-management behaviors. To improve engagement in self-management activities among people after stroke, we added a 12-week interactive SMS text messaging component to iSMART to identify and set treatment goals and provide daily goal-specific tips, personalized reminders, and motivational messaging, as well as a weekly goal check-in to promote the self-monitoring of progress. As people after stroke often experience restricted life participation, we also incorporated weekly coaching sessions into the existing iSMART framework to incorporate live weekly health coaching sessions to assess progress toward goals as well as identify barriers to engagement in self-management behaviors and strategies to address barriers, including changing or adjusting goal type or difficulty.

We hypothesized that engaging clinicians and researchers in the iSMART development and adaptation process using the IM framework would benefit the development process by incorporating the perspective of real-world clinical practice and improving the understanding of technology-supported self-management approaches, thereby increasing the uptake of such tools in the clinical management of stroke recovery. Here, we describe the use of IM to systematically adapt evidence-based self-management approaches for digital delivery, incorporating the perspectives of patients, their caregivers, and clinicians who treat people after stroke. The process and basic intervention framework may be relevant to other chronic conditions where self-management is key to promoting recovery and maintaining functional gains achieved during recovery.

**Methods**

**Overview of iSMART’s Intervention Architecture and Theoretical Foundations**

We first present an overview of iSMART and its theoretical foundations so that readers can better understand iSMART intervention. Behavioral determinants, mechanisms of action (MoAs), and screenshot examples are presented alongside the IM processes described later. iSMART is a 12-week digital intervention to improve skills to manage chronic conditions and support daily activity participation for people after stroke. iSMART consists of 3 components: a group psychoeducation program for skills training and practices, individual coaching, and SMS text messaging. The psychoeducation component of iSMART, which was adapted from IPASS, combines two theoretical approaches: (1) a chronic disease self-management program built on social cognitive or learning theory [28] and (2) community participation and environment management built on the person-environment-occupation-performance model [29]. iSMART’s psychoeducation component focuses on teaching 5 self-management skills (ie, problem-solving, accommodations, communication, decision-making, and emotion and symptom management) to manage chronic conditions and support daily activity participation in people after stroke.

The original IPASS program did not include a coaching component. Live health coaching is effective in chronic disease management and as such is commonly used across several chronic disease states as a tool for both treatment and prevention. Health coaching was incorporated into iSMART to promote collaborative goal setting and provides a mechanism for discussing and adjusting goal type and difficulty to promote self-efficacy for self-management behaviors. The coaching component of iSMART was grounded in principles of behavioral activation [30]. We adapted the Brief Behavioral Activation Treatment for Depression-Revised [31] manual for iSMART, in particular adapting the methods for identifying treatment goals based on valued life areas and activities, mechanisms for monitoring progress, and potential barriers to engagement or success, as well as proactively developing strategies to achieve goals based on these inputs. Both psychoeducation and coaching sessions were delivered weekly via videoconferencing for the first 7 weeks. Table 1 shows the content of the psychoeducation and coaching sessions.

To improve user engagement, we added the SMS text messaging component to iSMART to further support self-management behaviors. The messaging framework used in this study was adapted from prior studies with effectiveness demonstrated in hospital workers [32,33] and adults with severe mental illness [34]. As a first step toward adapting the platform for improving engagement in, and uptake of, self-management behaviors in people after stroke, content was adapted. Specifically, SMS text messaging content for goal reminders, self-management tips, motivational messaging, and monitoring for psychological distress (such as depressed mood) was customized to meet the needs of people after stroke and incorporated into the existing SMS text messaging framework. Figure 1 shows screenshots of different types of SMS text messages. These messages were sent following a predefined weekly schedule (Figure 2).
Table 1. Content of individual coaching and psychoeducation group sessions.

<table>
<thead>
<tr>
<th>Session</th>
<th>Individual behavioral activation coaching</th>
<th>Psychoeducation group</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Discussion of stroke</td>
<td>Introduction to the workshop</td>
</tr>
<tr>
<td></td>
<td>Introduction to treatment rationale</td>
<td>Introduction of group members</td>
</tr>
<tr>
<td></td>
<td>Introduction to daily monitoring</td>
<td>Things that limit or enable what you want to do: inside you</td>
</tr>
<tr>
<td></td>
<td>Important points about the structure of the treatment</td>
<td>Things that limit or enable what you want to do: outside you</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Problem-solving</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Figuring out a home activity to work on</td>
</tr>
<tr>
<td>2</td>
<td>Daily monitoring: review assignment</td>
<td>Planning what you want to do</td>
</tr>
<tr>
<td></td>
<td>Treatment rationale: review</td>
<td>Taking apart the activity</td>
</tr>
<tr>
<td></td>
<td>Complete life areas, values, and activities inventory</td>
<td>Problem-solving your home activity</td>
</tr>
<tr>
<td></td>
<td>Complete activity selection and ranking</td>
<td>Taking apart the work, volunteer, or other service position</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Closing</td>
</tr>
<tr>
<td>3</td>
<td>Daily monitoring: review assignment</td>
<td>Figuring out important community activities</td>
</tr>
<tr>
<td></td>
<td>Life areas, values, and activities inventory: review assignment</td>
<td>Defining community participation</td>
</tr>
<tr>
<td></td>
<td>Daily monitoring with activity planning and action planning</td>
<td>Problem-solving meaningful work</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Identifying and problem-solving a community activity</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Requesting accommodations in the community</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Closing</td>
</tr>
<tr>
<td>4</td>
<td>Daily monitoring with activity planning and action planning: review assignment</td>
<td>Feedback on the action plan developed with the coach earlier in the week</td>
</tr>
<tr>
<td></td>
<td>Daily monitoring with activity planning and action planning for the upcoming week</td>
<td>Identifying meaningful work, volunteer, or service positions</td>
</tr>
<tr>
<td></td>
<td>Contracts</td>
<td>Requesting reasonable accommodations at work</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Decision-making</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Closing</td>
</tr>
<tr>
<td>5</td>
<td>Daily monitoring with activity planning and action planning: review assignment</td>
<td>Feedback on the action plan developed with the coach earlier in the week</td>
</tr>
<tr>
<td></td>
<td>Contracts: review assignment</td>
<td>Making informed treatment decisions</td>
</tr>
<tr>
<td></td>
<td>Daily monitoring with activity planning and action planning for the upcoming week</td>
<td>Communication skills</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Communicating with family and friends</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Working with your health care professionals and health care organization</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Closing</td>
</tr>
<tr>
<td>6</td>
<td>Daily monitoring with activity planning and action planning: review assignment</td>
<td>Feedback on the action plan developed with the coach earlier in the week</td>
</tr>
<tr>
<td></td>
<td>Daily monitoring with activity planning and action planning for the upcoming week</td>
<td>Dealing with depression</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dealing with difficult emotions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Positive thinking</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Closing</td>
</tr>
<tr>
<td>7</td>
<td>Daily monitoring with activity planning and action planning: review assignment</td>
<td>Feedback on the action plan developed with the coach earlier in the week</td>
</tr>
<tr>
<td></td>
<td>Daily monitoring with activity planning and action planning for the upcoming week</td>
<td>Relaxation techniques</td>
</tr>
<tr>
<td></td>
<td>Life areas, values, and activities inventory: concept review and edit</td>
<td>Looking back and planning for future</td>
</tr>
<tr>
<td></td>
<td>Activity selection and ranking: concept review and edit</td>
<td>Closing and wrap-up celebration</td>
</tr>
<tr>
<td></td>
<td>Contracts: concept review and edit</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Preparing for the end of treatment</td>
<td></td>
</tr>
</tbody>
</table>
Figure 1. Screenshots of various types of SMS text messages. (A) goal reminder; (B): goal check-in; (C) self-management tip; (D) ecological needs assessment; (E) general motivation; and (F) mood check-in.

Figure 2. SMS text messaging schedule.

During the first 3 weeks, the coach worked with each participant to explore their valued life areas and activities and identify the treatment goals. A total of 25 predefined goals were programmed into iSMART. These goals target the application of self-management skills to improve participation in major life areas (ie, daily responsibilities, relationships, interests and recreation, education and career, and mind, body, and spirituality) derived from the behavioral activation treatment manual [31]. The coach entered up to 3 selected goals into the dashboard. Participants received the first SMS message at week 3, and for 10 consecutive weeks (ie, from week 3 to week 12), they received SMS messages every week. Multimedia Appendix 1 shows the screenshots of the iSMART dashboard. The dashboard can show the performance metrics of the study participants. It can also allow the coach to add, remove, or adjust goals and goal levels based on each survivor’s dynamic needs and preferences. The platform can automatically adjust self-management tips matched to the new goals sent to each participant.
survivor. In addition, the dashboard provides a secure chat feature that allows direct 2-way SMS text message communications between participants and the coach. All data collected via the apps, including SMS text messages sent or received through the secure chat function, were stored on Health Insurance Portability and Accountability Act–compliant servers behind the study institution’s firewall.

Ethics Approval

The institutional review board of Washington University (202004137) and Northwestern University (STU00215743) approved this study.

IM Approach

IM is a 6-step process often applied to guide the development, implementation, and evaluation of behavior change interventions [21,22]. It is a framework used to bridge the gap between theories and practice. Each step consists of several tasks, which, once completed, inform the next step. This rigorous framework ensures that the developed intervention can be implemented in real-world settings. The development process for iSMART was guided by the first four steps of IM, as illustrated in Figure 3, including (1) conducting a needs assessment, (2) specifying behavioral determinants and performance objectives, (3) applying theories and designing the intervention, and (4) developing and refining the intervention. Steps 5 (adoption and implementation plan) and 6 (evaluation plan) are being conducted in a separate clinical trial.

Figure 3. Intervention mapping steps. BATD-R: Brief Behavioral Activation Treatment for Depression-Revised; BCT: behavior change technique; COM-B: capability, opportunity, motivation, and behavior; IPASS: Improving Participation After Stroke Self-Management; iSMART: interactive Self-Management Augmented by Rehabilitation Technologies.
**Results**

In this section, we outline the tasks used in each IM step and describe how these tasks were completed, followed by reporting key findings of each step.

**Step 1: Needs Assessment**

**Overview**

Step 1 consisted of a detailed multimethod assessment of the needs of people after stroke to manage chronic conditions and participate in meaningful activities. This first IM step includes three tasks: (1) forming a planning group, (2) conducting a telephone survey, and (3) conducting a systematic review. The key findings of these tasks are provided in the following subsections.

**Planning Group**

The planning group (n=6) comprised stakeholders with expertise in stroke care, self-management, and technology development, including an occupational therapist in inpatient stroke rehabilitation, an occupational therapist in community rehabilitation, a director at a rehabilitation clinic, a director working at a technology company, a PhD-level nurse researcher with expertise in self-management, and a person living with chronic conditions after stroke. They had a wide range of experience in their content areas (mean 7.90, SD 5.44 years). They worked with the research team to adapt, test, and provide feedback on iSMART’s content, format, and perceived feasibility.

**Telephone Survey**

We conducted a telephone survey regarding mobile technology–supported health services with people after stroke (n=125). Participants were recruited from a stroke registry of patients admitted to a stroke center in the Midwestern United States. Our research assistants conducted telephone surveys with people after stroke about their use of mobile devices and preferences for technology-enabled services and formats for stroke rehabilitation. Survey questions were developed based on the Pew Research Center’s survey report regarding technology device ownership [35] and smartphone use [36], as well as a study of mobile technology services among persons with mental illness [37]. Multimedia Appendix 2 shows the survey questions. Please refer to our prior study [38] for participants’ eligibility and other details.

Of the 125 study participants, 79 (63.2%) were smartphone users (mean age 60.5, SD 13.0 years), and 46 (36.8%) were non–smartphone users (mean age 70.5, SD 10.9 years). Of the 79 smartphone users, 39 (49%) were men, and 44 (56%) were White; whereas, of the 46 non–smartphone users, 24 (52%) were men, and 33 (72%) were White. The top 5 desired services rated by smartphone users were appointment or activity scheduling, symptom management, stroke prevention, cognitive enhancement training, and access to stroke care resources. Non–smartphone users also rated these 5 services as their most desired services, plus socialization, as their most desire services. Smartphone users also reported that making video or voice calls, SMS text messaging, and surfing the internet were the top 3 most common functions associated with their mobile device use over the last 12 months. On the basis of these findings, we decided to focus the iSMART content on helping participants to identify and schedule their valued activities into their daily routine; remind them about, and monitor the progression of, goal completion; monitor and manage symptoms; and provide resources and knowledge associated with self-management and secondary stroke prevention. For the iSMART’s delivery format, we used videoconferencing to run the psychoeducation and coaching sessions and SMS text messaging (with a clinician-facing dashboard to customize SMS text messages to participants) to support self-management behaviors.

**Literature Review and Meta-Analysis**

We conducted a literature review to identify personal (medical and behavioral) and environmental factors associated with the inability to manage chronic conditions and symptoms leading to restricted participation in meaningful activities. We summarized and produced a logic model (Figure 4) to link personal (medical and behavioral) determinants and environmental determinants, leading to the target health and participation problems.
We also conducted a meta-analysis study of digital self-management interventions in people with neurological disorders. The review aimed to identify the theory, outcomes, and optimal mode of intervention delivery for developing iSMART. A prior manuscript describes our detailed search strategy, study selection, data extraction, quality appraisal, and statistical analyses [19]. We found that interventions based on cognitive behavioral theory were effective in reducing depression, anxiety, and fatigue and enhancing self-efficacy. By contrast, interventions based on social cognitive theory were effective in reducing depression only. In addition, digital self-management interventions that incorporated live health coaching or support from a health professional were found to be more effective than fully digital self-guided interventions. Thus, we incorporated live health coaching into the digital intervention framework, facilitated through scheduled videoconferences and both scheduled and ad hoc bidirectional SMS text messages.

**Step 2: Behavioral Determinants and Performance Objectives**

Informed by the findings from the needs assessment, we used the capability, opportunity, motivation, and behavior model [39] to guide the identification of key behavioral determinants of self-management. We then derived performance objectives that mapped to each behavioral determinant. Performance objectives are actions taken by people after stroke to achieve behavioral determinants.

The study team, which consisted of experts in digital health intervention development, adaptation, and testing, identified seven behavioral determinants most likely to affect the treatment goal and outcomes: (1) knowledge, (2) behavioral regulation, (3) skills, (4) self-efficacy, (5) motivation, (6) negative and positive affect, and (7) social support and environmental support. We further derived performance objectives mapped to each behavioral determinant (Textbox 1).

<table>
<thead>
<tr>
<th>Knowledge</th>
<th>Behavioral regulation</th>
<th>Skills</th>
<th>Self-efficacy</th>
<th>Motivation</th>
<th>Negative and positive affect</th>
<th>Social support and environmental support</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Learn about stroke consequences, benefits of the intervention, factors, and resources available to support self-management behaviors and activity participation</td>
<td>• Monitor progress in increasing self-management behaviors and meaningful activity participation</td>
<td>• Learn self-management skills and strategies</td>
<td>• Increase confidence to use self-management strategies effectively</td>
<td>• Increase engagement in planned goals and activities and engage with the digital program</td>
<td>• Cope with challenges associated with self-management behaviors and engagement in planned activities</td>
<td>• Provide resources to peers and increase access to resources</td>
</tr>
</tbody>
</table>

Step 3: Intervention Design

We developed a framework to identify target behaviors and outcomes, considering the MoAs most likely to affect the selected behavioral determinants. MoAs are constructs that can be individual characteristics (eg, knowledge and attitudes) or contextual characteristics (eg, social support and environmental resources) to mediate intervention effects [40]. Next, we used the linkage table published by Carey et al [40] to match the behavior change techniques (BCTs) to each of the MoAs. Subsequently, the planning group members developed by consensus a set of practical empirically supported approaches to address the selected BCTs.

**Figure 5** outlines the theoretical framework used to inform the design, additional content, and functionality of iSMART. We identified 14 MoAs affecting behaviors (ie, adherence to the intervention, chronic condition and symptom management, and engagement in valued activities), ultimately leading to improved outcomes (ie, physical health, psychosocial health, and activity participation) for people with disabilities. **Multimedia Appendix 3** describes the MoAs, BCTs, and practical application ideas used for iSMART development.
Step 4: Intervention Development

Overview
We developed all iSMART components based on the findings from steps 1 to 3. For the psychoeducation component, we incorporated the practical application ideas to adapt the original IPASS content into iSMART. For the coaching component, we incorporated the Brief Behavioral Activation Treatment for Depression-Revised manual [31] into iSMART. For the SMS text messaging component, we adapted an SMS platform used in previous studies [32-34], including a clinician-facing dashboard and SMS text messaging libraries. Although the iSMART SMS text messaging followed the same weekly messaging schedule as the parent platform, we developed new SMS text message content customized to improve skills in managing chronic conditions and support participation in daily life and community activities for people after stroke. Next, the planning group members reviewed all materials, including the iSMART manual, treatment forms, and SMS text messages. A health literacy specialist also reviewed these materials to improve health literacy elements and readability. We revised the treatment materials and liaised with a digital health start-up to refine the digital platform. Afterward, each of the planning group members trialed the refined program for 2 to 3 weeks and completed three 4-item implementation outcome measures [41]—the Acceptability of Intervention Measure (AIM), the Intervention Appropriateness Measure (IAM), and the Feasibility of Intervention Measure (FIM)—to assess the preimplementation feasibility of iSMART. Items were rated on a 5-point scale, ranging from 1=completely disagree to 5=completely agree. Benchmarks for high acceptability, appropriateness, and feasibility of the study intervention are mean scores of 4 out of 5 on the AIM, IAM, and FIM. These measures had strong structural validity with Cronbach $\alpha$ values of .85 for the AIM, .91 for the IAM, and .89 for the FIM as well as strong test-retest reliability with Cronbach $\alpha$ values of .83 for the AIM, .87 for the IAM, and .88 for the FIM [41]. The planning group members also provided written feedback for further improvement.

We have reported the intervention architecture and theoretical foundations of iSMART in the Methods section. We focus on reporting major changes to iSMART in the Results section.

Major Changes to iSMART
The planning group members suggested minor but important changes to the order in which content is delivered to improve iSMART. First, they suggested reorganizing the psychoeducation content. Specifically, they recommended moving the problem-solving module to the first session so that participants could learn these foundational skills before attempting more challenging skills, such as requesting accommodations. They also suggested that the dealing with depression content should be delivered before the positive...
thinking and dealing with difficult emotions content. It was hypothesized that this reorganized learning sequence would avoid people experiencing depressive symptoms to find strategies learned via positive thinking and dealing with difficult emotions content less effective. They also recommended that home-based content be delivered before community-based content, allowing participants to practice their skills in a more familiar setting. Second, the planning group members advised dividing the 15-minute break into one 5-minute break and one 10-minute break during group sessions to optimize participants’ attention and engagement. Third, to reduce participant burden, the planning group members recommended that if the coach chose to implement iSMART in a one-on-one format, the coach could select relevant self-management psychoeducation modules based on the participant’s needs instead of teaching all 5 modules.

Finally, the planning group members suggested adjusting the time commitment for completing the different treatment forms or homework. Instead of asking participants to report hourly activities 7 days a week, they proposed that the coach and participant select 2 days of the week for reporting activities that best represent their regular rescheduling within the first 3 coaching sessions. In addition, the planning group members voiced concern that participant comfort and familiarity with mobile health technology and SMS text messaging would potentially limit engagement and affect the intervention’s overall effect. Thus, we developed simplified educational content to support the use of common digital tools (eg, videoconference tool) on mobile devices.

Preimplementation Feasibility of iSMART

The planning group members scored the acceptability, appropriateness, and feasibility (on a scale of 1 to 5, with 5 being most favorable) of the iSMART program as high, with mean scores of 4.63 (SD 0.38), 4.63 (SD 0.38), and 4.58 (SD 0.34), respectively. These findings suggest that our iSMART is an acceptable, appropriate, and feasible program.

Discussion

Principal Findings

In this study, we present a use case of the application of IM and BCTs to adapt and develop a digital intervention to improve the self-management of chronic conditions and daily activity participation in people after stroke. Although digital behavioral interventions are becoming increasingly popular, more research is needed to guide the process of translating evidence, theories, existing interventions, and user feedback for use in developing or adapting digital behavioral interventions. This study can serve as a design blueprint for researchers aiming to digitize self-management or other behavioral programs to improve intervention access, engagement, and effectiveness. This study has applied multiple empirically supported theories and stakeholder input to inform intervention development (using the IM framework to organize intervention inputs), produce an overarching logic model, and identify MoAs and BCTs to guide intervention development.

Prior Works and Study Implications

The technology-supported delivery of behavioral interventions holds promise for improving the precision of behavioral interventions by allowing the intervention to be tailored to the user and adapted over the course of the intervention as the user makes progress [42]. The iSMART platform uses an innovative architecture that facilitates personalized interaction and accessible resources provided to the user on demand.

As in the case of other chronic conditions, people after stroke often face challenges with limited access to needed health services [43]. iSMART uses SMS text messaging and videoconferencing, which can provide an alternative delivery solution to those who would not otherwise have access to these essential services [14]. The iSMART intervention can be expanded to reach many mobile phone users at a low cost and address clinical barriers to access. As of 2021, about 97% of American adults own a mobile phone, and 85% own a smartphone [44]. In addition, 73% of mobile phone owners use SMS text messaging on their mobile phones, and these SMS text messaging users send or receive an average of 41.5 messages per day [45]. Those with lower income (ie, <US $30,000 per year) and education (ie, <high school) mainly rely on their mobile phones for web-based access [46]. Digital interventions can overcome traditional barriers to patients receiving rehabilitation only at hospitals, such as inconvenience, geographic isolation, and financial burden [47].

Limitations and Future Directions

Because of the funding constraint, the planning group is small, and only 1 person with stroke has been included. Recognizing that patient and public involvement is essential to strengthen and improve the quality of the tool and make it more relevant [48], future research should involve more people after stroke and their caregivers in the development process.

The planning group makeup is such that it does not represent individuals from populations that may be at higher risk for chronic health conditions associated with stroke risk, including populations with low literacy; those who are socioeconomically disadvantaged and may have limited access to, or experience using, technology; and underrepresented communities, such as racial, ethnic, and gender minority groups. Adaptations to the iSMART intervention may not fully cover the unique needs of these vulnerable populations. Future steps include engaging a more heterogeneous population of people after stroke or persons with other disabilities in user-centered design activities that may increase the adoption and sustainable implementation of iSMART to a broader population of end users. The use of IM for intervention development is a resource-intensive process, consuming substantial efforts and resources. Nonetheless, we found the IM to be a valuable framework for guiding interventions for the target population.

Conclusions

This paper demonstrates the use of IM and BCTs to support the adaptation and development of an intervention designed to promote poststroke self-management skills to improve the management of chronic conditions and promote daily life participation. The rigorous process results in higher transparency
in understanding treatment mechanisms and allows replications for designing other complex interventions. This paper can be a valuable blueprint for developing digital interventions for self-management in different conditions.

Acknowledgments
The authors would like to thank the planning group members for their time and effort in providing feedback related to the intervention development. The authors would also like to thank graduate students, staff, and faculty members at Washington University for their research assistantship, in-kind resource sharing, mentorship, and involvement in different aspects of the research. The contents of this publication and the writing effort of the first author were supported by grants from the American Occupational Therapy Foundation (AOTFIRG20Wong) and the National Center for Medical Rehabilitation Research under the Eunice Kennedy Shriver National Institute of Child Health and Human Development (K01HD095388). Research reported in this publication was also supported by the National Institute of Mental Health (R34 MH118395) and the Washington University Mobile Health Research Core, part of the Institute of Clinical and Translational Sciences, funded by the National Center for Advancing Translational Sciences (UL1TR002345). The content is solely the responsibility of the authors and does not necessarily represent the official view of the funding agencies.

Data Availability
The data generated and analyzed during this study cannot be sufficiently deidentified and, therefore, cannot be made publicly available owing to ethical considerations. Deidentified data can be made available, for further research, by the corresponding author on reasonable request.

Authors’ Contributions
AWKW contributed to the study’s conception and design, materials preparation, data management, and analysis. AWKW and MWMF wrote the first draft of the manuscript. AWKW, MWMF, EGSM, CLM, SIL, GEN, OD, SET, KJK, and DCM commented on previous versions and read and approved the final manuscript.

Conflicts of Interest
GEN has received research support from the National Institutes of Health (NIH), the Health Resources and Services Administration, the Barnes Jewish Hospital Foundation, the Washington University McDonnell Center for Systems Neuroscience and Mallinckrodt Institute of Radiology, and Usona Institute (drug only). She has served as a consultant for Alkermes plc; CarelonRx; Otsuka Pharmaceutical Co., Ltd; and Sunovion Pharmaceutical Co., Ltd. DCM reports research support from the NIH. He has served as a consultant for Otsuka Pharmaceutical Co., Ltd; Optum Behavioral Health; Centerstone Research Institute; and OneMind Foundation. He receives royalties from Oxford Press and has an ownership interest in Adaptive Health. SIL reports research support from the NIH. MWMF has served as an independent contractor for Isaac Ray Forensic Group and Michigan Avenue Neuropsychologists. CLM has an ownership interest in Infinite Arms. He reports subcontracts from the NIH and VA Headache Centers of Excellence. AWKW reports research support from the NIH; the National Institute on Disability, Independent living, and Rehabilitation Research; and the Craig H. Neilsen Foundation. All other authors declare no other conflicts of interest.

Multimedia Appendix 1
Screenshots of the interactive Self-Management Augmented by Rehabilitation Technologies (iSMART) dashboard.
[PDF File (Adobe PDF File), 687 KB - humanfactors_v10i1e45099_app1.pdf ]

Multimedia Appendix 2
Survey questions.
[PDF File (Adobe PDF File), 174 KB - humanfactors_v10i1e45099_app2.pdf ]

Multimedia Appendix 3
Mechanisms of action linked to behavior change techniques and the design of practical applications.
[PDF File (Adobe PDF File), 156 KB - humanfactors_v10i1e45099_app3.pdf ]

References


Abbreviations

AIM: Acceptability of Intervention Measure
BCT: behavior change technique
FIM: Feasibility of Intervention Measure
IAM: Intervention Appropriateness Measure
IM: intervention mapping
IPASS: Improving Participation After Stroke Self-Management
iSMART: interactive Self-Management Augmented by Rehabilitation Technologies
MoA: mechanism of action

©Alex W K Wong, Mandy W M Fong, Elizabeth G S Munsell, Christopher L Metts, Sunghoon I Lee, Ginger E Nicol, Olivia DePaul, Stephanie E Tomazin, Katherine J Kaufman, David C Mohr. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 24.07.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
A Tailored mHealth App for Improving Health and Well-Being Behavioral Transformation in UK Police Workers: Usability Testing via a Mixed Methods Study

Richa Mehra¹, BPT, PGDHHM, MSc; Andy Pulman², MA, PhD; Huseyin Dogan¹, BSc (Hons), PG Cert, EngD, MSc; Jane Murphy³, BSc (Hons), PhD; Fiona Bitters³, BSc (Hons), MSt

¹Bournemouth University, Faculty of Science and Technology, Bournemouth, United Kingdom
²Bournemouth University, Faculty of Health and Social Sciences, Bournemouth, United Kingdom
³Hampshire Constabulary, Southampton, United Kingdom

Corresponding Author:
Andy Pulman, MA, PhD
Bournemouth University, Faculty of Health and Social Sciences
Bournemouth Gateway Building
Lansdowne Campus, St. Pauls Lane, Bournemouth
Bournemouth, BH8 8GP
United Kingdom
Phone: 44 1202 961047
Email: apulman@bournemouth.ac.uk

Abstract

Background: When considering the policing environment of 2022, many roles previously in the domain of warranted officers (police officer) are now performed by nonwarranted police staff equivalents. These police staff roles have expanded rapidly into other areas such as investigations, custody, and contact management, which were traditionally seen as police officer functions and put staff under some of the same stresses as police officers. A UK police force requested help in investigating technologies that could be used to improve health and well-being for both officers and staff.

Objective: The aim of this study was to create a health and well-being app for police officers and staff, which considered the unique requirements of the users throughout the designing, building, prototyping, and testing stages.

Methods: This study involved quantitative approaches (demographic web-based survey questions and the System Usability Scale) and qualitative approaches (open web-based survey questions and semistructured interviews). Unsupervised usability testing of a prototype app was undertaken by members (N=48) of the commissioning client using their smartphones. After completing a preregistration application for screening purposes, participants downloaded a trial version of the app. Then, they completed a web-based questionnaire after testing the app for 10 days. A subsample of participants (9/48, 19%) was interviewed. Deductive thematic analysis was undertaken to identify key themes and subthemes.

Results: Data collected during usability testing concerned the 6 domains of the app—food and diet, activity, fluid intake, sleep, good mental health, and financial well-being—and informed the creation of improved design during prototyping. Some usability and design issues and suggestions for improvements were also addressed and implemented—including shift management and catch-up cards—during this cycle of development.

Conclusions: This study highlights the importance of coparticipation with officers and staff across the entire development cycle, to coproduce a human-centered design methodology to enable the development of a considered and user-centered solution. It demonstrates the need for producing a multifunctional tool rather than focusing purely on an individual element for this user group. It also highlights how linking and being able to track optional, personalized elements of health data against one another, cross-referenced to individual shift patterns, might help to inform and provide users with a chance for reflection and therefore influence behavior change.

(JMIR Hum Factors 2023;10:e42912) doi:10.2196/42912

https://humanfactors.jmir.org/2023/1/e42912
KEYWORDS
nutrition; activity; behavior change; telemedicine; mobile health; police; lifestyle management; management; usability testing; design; build; prototype; testing; survey; interview; development; user center; officer; law enforcement; cop; detective; policeman; policing; mobile phone

Introduction
Police officers can experience difficulties when managing health and well-being as a result of working long, unsocial hours in a highly pressurized environment [1,2]. Those working in the police face different health challenges from those in the general population, as their job might feature unusual working hours and alternating shift patterns [3]. Police officers have a great chance of being overweight and obese and are at risk of long-term conditions such as cardiovascular disease [4] and cancer [5] compared with the average population. Risk factors are heightened by working in a highly stressful environment, increasing the likelihood or the severity of these issues [6,7]. Compared with the average person, an officer is more likely to experience stress owing to exposure to dangerous situations and traumatic events [8]. Stress can have multiple knock-on effects including insomnia, fatigue, and poor concentration—all of which make performing the job harder than it already is and add to the original problem [9].

When considering the policing environment of 2022, many roles previously in the domain of warranted officers are now performed by nonwarranted police staff equivalents. These police staff roles have expanded rapidly into other areas such as investigations, custody, and contact management, which were traditionally seen as police officer functions and put staff under some of the same stresses as police officers. Officers and staff can also be affected by mental health conditions such as posttraumatic stress disorder, depression, and alcohol abuse [10]. Physical risks are also an issue with increased danger of long-term health problems such as back pain [11,12]. Studies during the COVID-19 pandemic also suggest that stress levels of policing have increased during this period and have affected officer resiliency [13,14].

Although some risk factors relating to health and an individual’s risk of certain health issues are nonmodifiable such as age and genetics, modifiable lifestyle factors can be self-managed by individuals to reduce the risk. Currently, there is a large number of health and fitness solutions available on multiple platforms, but none are tailored to specific issues that the police force face [15]. Data highlighted that a multifunctional tool would be more beneficial than focusing on a single element. Key features and 4 domains were identified for initial app coverage. In order of importance—prioritized by participant response—these were: food and diet (76/213, 35.6%), activity (68/213, 31.9%), sleep (27/213, 12.6%), and fluid intake (27/213, 12.6%). Participants also identified a need for the new app to consider that a user was on shift—this is important because many issues and problems with elements of their health and well-being involved shift work. For example, shift work and fatigue have been shown to interfere with sleep and impair cognitive function [16,17].

Secondary Design Cycle
Initial requirements were categorized using the MoSCoW (Must Have, Should Have, Could Have, Won’t Have) framework [18], with findings from the web-based questionnaire and client meetings informing initial draft requirements. Paper designs were sketched leading to the creation of low-fidelity wireframes [15]. These were shown to the client for feedback before high-fidelity designs were created. These were then shown to interviewees during semistructured interviews (n=10), to gather feedback about requirements and preferences for the app. On the basis of their feedback, the second set of design prototypes was created. A good mental health section was added as the fifth domain, including the ability to complete an optional mood diary and track and set alcohol goals. Given the emerging evidence about the impacts of COVID-19 on police staff’s mental well-being before [10,19] and during the pandemic [13,14], the suggested expansion to incorporate a section regarding mental health and well-being seemed to be a valuable addition, especially in light of potential challenges broaching these topics within a culture where discussing mental health difficulties have sometimes been viewed as an undesirable discussion topic [20]. Considerations and concerns of financial nature [21-23] inspired the inclusion of the sixth domain, to cover holistic wellness. It was highlighted that information regarding financial planning, pension policies, and budget and saving advice would be helpful and give control to those who were nearing retirement. Overall wellness is something achieved by taking care of mental, physical, and financial well-being. The suggestion to include financial well-being was incorporated late into the second development cycle.

Testing Version
Further development using the revised high-fidelity designs as an initial foundation was undertaken with an external developer to produce a prototype for pilot testing. The app—created in Android and Apple iPhone Operating System—was developed using an agile approach. Coding was conducted between October 2021 and April 2022. Several revised versions were developed to include all MoSCoW requirements identified as “must” and gather feedback and information from the user base. Survey data were gathered from 213 participants of the UK commissioning force [15]. Data highlighted that a multifunctional tool would be more beneficial than focusing on a single element. Key features and 4 domains were identified for initial app coverage. In order of importance—prioritized by participant response—these were: food and diet (76/213, 35.6%), activity (68/213, 31.9%), sleep (27/213, 12.6%), and fluid intake (27/213, 12.6%). Participants also identified a need for the new app to consider that a user was on shift—this is important because many issues and problems with elements of their health and well-being involved shift work. For example, shift work and fatigue have been shown to interfere with sleep and impair cognitive function [16,17].
“should” have [18]. The process of developing the app to this point was highly iterative and agile.

**Study Design**

The purpose of this study was to create a health and well-being app for police officers and staff, which considered the unique requirements of the users throughout the designing, building, prototyping, and testing stages and helped the researchers understand the users and their requirements. The study involved quantitative approaches (demographic web-based survey questions and the System Usability Scale [SUS] [24]) and qualitative approaches (open web-based survey questions and semistructured interviews). We used a human-centered approach to optimize the understanding and accommodate the perspectives of potential users [25]. The methods chosen were considered with the participants in mind—those that could be completed in 1 session, where time could be set aside when they were free. The methods used in this study acknowledged and catered to the logistical and operational pressures that the participants might be under, while allowing them to provide the project team with detailed feedback.

**Participants, Recruitment, and Consent**

Participants were recruited via a gatekeeper from the commissioning client. The gatekeeper’s role was to initiate communication between the researcher and police officers and staff who wished to participate—via distribution of study literature—without compromising anonymity or affecting the veracity of web-based responses. Participation was voluntary, and participants remained anonymous to both the gatekeeper and the organization. Consent for completing the web-based survey was requested before allowing participants to proceed. Before each interview began, the researcher answered any questions they had, checked if they were willing to be audio recorded, and explained the consent process before recording began.

**Procedures and Measures**

This study involved unsupervised usability testing of a prototype app by members of the commissioning client using their smartphones. The version of the app supplied for testing contained draft versions of all 6 sectional domains previously identified—food and diet, activity, sleep, fluid intake, good mental health, and financial well-being—accessible from the home screen (Figure 1). An additional feature provided the opportunity to add forthcoming shift pattern details to the app, which could then be cross-referenced with other parts of the app. The testing version concentrated on implementing basic operational functionality, which could then be expanded or altered as required based on user feedback.

**Figure 1.** Screenshot of home screen.
Initially, an email was sent to police officers and staff describing the study and asking about their interest. After completing a preregistration application for screening purposes, participants were emailed a link to allow them to download the trial app. A task guide was also sent, giving them 10 days of time for testing. The guide suggested task scenarios of different complexity levels and covered the core functionalities of the app: shift manager, fluid intake, sleep, mood diary, alcohol consumption tracker, financial well-being, and activity. Participants were then asked to complete a web-based questionnaire after 10 days of testing. Demographic details, information about use of the app, likes and dislikes, and feedback about specific features were collected through open-ended questions. The questionnaire was completed by 48 participants.

A subsample of participants (9/48, 19%) who provided consent to a follow-up telephone interview via a questionnaire subsection were interviewed after they had completed the given tasks. Topic guides were used for the semistructured interviews, to ensure that areas of interest (such as aspects related to delivery mode and format) were covered, while still allowing flexibility. Participants were asked several follow-up questions about the content, design, and functionalities of the app and to rate their experience. Throughout the session, the interviewer encouraged the participants to speak aloud about their actions, which helped to understand the emotions of the user while using the functionalities of the app. They were encouraged to share the issues faced while performing a task or give suggestions for improving the design. Interviews were audio recorded and transcribed verbatim. Notetaking was also used to collect data during the usability sessions.

Sample Size
As the intended audience for the final product was a UK police force, the work force value supplied by the client organization was rounded to the nearest 1000 and used to calculate an ideal sample size. The value used for the work force was 5000 [26]. The actual number of participants was 48, which was above the lower threshold for an acceptable number for the sample size. With this number of participants and the estimated population size, there is a 95% confidence level, with a final margin of error of 14%. This sample size was considered to be adequate because of restrictions in accessing participants—for downloading and testing the app, completing the survey, and possibly participating in an interview—owing to logistical and operational pressures.

Analysis
Recordings were transcribed and thematically analyzed using a deductive approach that focused on the domains covered in the topic guide (focusing on design, functionalities, and content). A generic qualitative approach to thematic analysis was used [27], with interresearcher interpretation.

Ethics Approval
Ethics approval was obtained from Bournemouth University (39100).

Results
Participant Characteristics
We recruited 33% (16/48) male and 67% (32/48) female participants. The “other” or “prefer not to say” option was also included within the survey question—no responses were received. (Table 1).
### Table 1. Self-reported descriptions about participants (N=48).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>16 (33)</td>
</tr>
<tr>
<td>Female</td>
<td>32 (67)</td>
</tr>
<tr>
<td><strong>Age group (years), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;18</td>
<td>0 (0)</td>
</tr>
<tr>
<td>18-30</td>
<td>4 (8)</td>
</tr>
<tr>
<td>31-40</td>
<td>20 (42)</td>
</tr>
<tr>
<td>41-50</td>
<td>15 (31)</td>
</tr>
<tr>
<td>51-60</td>
<td>8 (17)</td>
</tr>
<tr>
<td>61-70</td>
<td>1 (2)</td>
</tr>
<tr>
<td>&gt;70</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Job role, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Police officer</td>
<td>20 (42)</td>
</tr>
<tr>
<td>Police worker</td>
<td>28 (58)</td>
</tr>
<tr>
<td><strong>Shift worker, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>40 (83)</td>
</tr>
<tr>
<td>No</td>
<td>8 (17)</td>
</tr>
<tr>
<td><strong>System Usability Scale</strong></td>
<td></td>
</tr>
<tr>
<td>Score, median (range)</td>
<td>73.8 (30-100)</td>
</tr>
<tr>
<td>Score, mean (SD; 95% CI)</td>
<td>70.9 (14.1; 66.9-74.9)</td>
</tr>
<tr>
<td><strong>Adjective rating (Sauro-Lewis), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>A or A+</td>
<td>11 (23)</td>
</tr>
<tr>
<td>B</td>
<td>13 (27)</td>
</tr>
<tr>
<td>C</td>
<td>12 (25)</td>
</tr>
<tr>
<td>D</td>
<td>5 (10)</td>
</tr>
<tr>
<td>E</td>
<td>7 (15)</td>
</tr>
</tbody>
</table>

**Summary Statistics for the SUS**

Summary statistics for the SUS scores are presented in Table 1. Frequencies of ratings for each of the 10 SUS items are shown in Figure 2. Overall, the median SUS score for the toolkit was 73.8 and the mean was 70.9 (SD 14.1; range 30-100; 95% CI 66.9-74.9). This equates to an adjective rating of “OK” [24] and a “C” (41st to 59th percentile range) on the Sauro-Lewis curved grading scale [28]. Most participants (23/48, 48%) thought they “would like to use this toolkit frequently” (the version of the prototype they tested; SUS question 1). Most (40/48, 83%) considered the toolkit “easy to use” (SUS question 3), with 88% (42/48) believing that most people would “learn to use it quickly” (SUS question 7). More than half (28/48, 58%) considered the toolkit’s functionality to be well integrated (SUS question 5).
**App Feedback—Domains**

Feedback from survey data and interviews—encompassing themes of design, functionality, and content—was organized into the previously mapped areas of the app. Suggestions made for additional functionality were considered and included any relevant feedback elicited during interviews.

**Food and Diet**

Police personnel wanted to have a tracker on the app to monitor the food that they consumed daily. Most of the users interviewed (6/9, 67%) found the supporting information under this section to be quite relevant and informative. During interviews, 22% (2/9) of the participants mentioned that they liked the detailed information provided about immunity support:

*One really good to see the immunity support page, not a lot of people know how different foods can affect you and stuff.*  
[Participant 527]

Additional suggestions were the ability to view a summary of how changing shifts could affect healthy food consumption.

**Activity**

The activity section (Figure 3) is currently incomplete without a fully developed tracker, and information provision is not going to help in maintaining good fitness levels in isolation. However, most participants (7/9, 78%) thought that the information provided was relevant, helpful, and comprehensive. They especially mentioned that “Cycle2Work”—a UK government tax exemption scheme, introduced to promote healthy journeys to work—and other sections were good to have on an app as they enabled easy access to all the required wellness material from home. Older participants appreciated the information displayed on posters and found it easy to read. During interviews, participants also mentioned that they would like to track all the physical activities they were doing, including linking the app to their step counter:

*Ability to linking with step counter. Idea is to have everything at one place.*  
[Participant 517]

Another suggestion was the ability to share information within teams to act as a motivator:

*It definitely doesn’t need to rival Strava or MyFitnessPal to be useful but in the activity section a means of tracking exercise [only basic - not pace, incline etc] to perhaps make teams or share stats within or across shifts. To set up a Contact Management leader board of which shift is taking the most steps for example.*  
[Participant 617]
Fluid Intake

Feedback regarding the fluid intake section (Figure 4) included the calendar feature associated with the trackers not working properly to input fluid and alcohol; it displayed units of fluid or alcohol drunk on the current day, even if the entry was made for previous days. Overall, 22% (2/9) of the interviewed participants did not find the fluid tracker to be user-friendly and requested a more detailed help section to guide the user through this section. Participants wanted to have reminders set for fluid intake and record old entries for those that might have been missed. However, catch-up cards (refer to the App Feedback—Other Features section) that give reminders to drink water should ideally direct users immediately to the fluid intake tracker for a smooth transition. During interviews, 44% (4/9) of the participants mentioned the desirability of viewing a summary:

So you were making entries every day, but if one day if you would like to see the pattern of your fluid intake, you should get the kind of graph or some. [Participant 527]
Sleep
Interviewed participants (8/9, 89%) described the information provided about sleep as useful and quite vast. There were some comments about the consistent formatting of the content, and users gave their preference in the web-based survey to style as type 1, which was not only easy on the eyes but also more user-friendly.

During interviews, 56% (5/9) of the participants recommended having the ability to log sleep:

...Load loads of nice information. Yeah, but it would have been nice to have been able to have as a log. [Participant 530]

Good Mental Health
Alcohol tracker, mood diary, employee assistance information, and recognizing stress were some of the features that the participants tested. Only some could test all the subsections during the testing period. Regarding the alcohol tracker, some participants (2/9, 22%) did not expect this tracker to be located under this section. For them, it made more sense if it was moved to the fluid intake section. The tracker offers the ability to track how many units of alcohol a user has drunk that day, and in addition to this feature, interviewed participants (4/9, 44%) wanted to see an overall summary in relation to their shift, to understand the alcohol consumption and success in meeting the goal.

The interviewed participants (4/9, 44%) appreciated the National Health Service (NHS) information on feelings and symptoms that can be common with mental health concerns, linked from within the app to recognize the level of stress [29], because for some job roles that required working on the front line, operating at unsocial hours and occasionally experiencing traumatic events increased stress levels. Low mood is an indicator of poor mental health [30], and it is possible to track and compare with the previous days using this feature. Participants mentioned that the mood diary was something they would like to use in the future, as it helped to track changes in mood daily and compare them with previous days. It was viewed by some as being extremely important to keep checking this aspect of their health:

And you can kind of almost pull off reports. Maybe it’s a PDF report or something that shows that actually they may be able to identify that on those two days after a set of three lates. They’re moods quite low or their fluid intake’s really bad or do you know what I mean? So that people may utilize the app more... [Participant 532]

However, there was also some apprehension noted in using the mood diary. Concerns were whether the data might be monitored by senior management and that they might be viewed differently as a result.
Financial Well-Being
Participants liked the amount and types of information included in this section:

I there was some really useful stuff on there that was sort of flicking through with them, particularly around sort of the money saving. I read a lot about the police mutual bit and that was really good. And again gave me more stuff to think about and it’s got all the links for stuff that I use anyway, like the HPL bit and the blue Light card section and all that sort of stuff. [Participant 517]

Having all the information easily available on the app was appreciated. Participants wanted to see the search function when the reading list was long and screen scrolling was required.

App Feedback—Other Features
Dashboard and Visualization
Survey respondents (27/48, 56%) appreciated the clean look of the app and the fact that the home page was divided into subheadings, which not only eased navigation but also made it easy to use:

...Layout was quite visually appealing so the subheadings. So, you kind of, you knew what you were gonna get when you went into those subheadings,

They emphasized that they would like to see the summary dashboard in a graphical form for each domain in correlation with the shifts attended. It is always helpful to see the progress made to maintain the motivation levels high, which is crucial during self-monitoring.

Shift Manager
The shift manager (Figure 5) received positive feedback. A few different iterations were trialed to make this feature more user-friendly. The most common officer and staff shifts were made available to users, so that they could choose the one that suited their pattern best. In addition, the user can make modifications and customize the shift patterns if there were any last-minute changes.

However, there were some challenges. Some participants (18/48, 38%) suggested that the ability to view and edit the entire pattern of the shift in one go would save them manual input time. They also suggested replicating the edited pattern for future use:

I couldn’t figure out a way to put in like a six week shift pattern and then repeat it. I could only figure out how to do it manually day by day. you could do like a six week Pan and start it on a specific date and then it repeats itself, then obviously I could do it that way and it’ll be a lot easier. [Participant 517]
**Goal Setting and Catch-Up Cards**

Participants mentioned that the goal-setting feature was useful when using it within the fluid intake section. There were suggestions to include this feature in other domains.

In the test version, a new functionality—catch-up cards (Figure 6)—was introduced in association with shift management improvements. These cards help to remind the users about certain tasks and popped up based on the shift chosen.

**Figure 6.** Screenshot of catch-up card - wellbeing walks.

Participants wanted to use catch-up cards as reminders to meet the set goals. The ability to change the frequency of catch-up cards on an individual basis was something that participants wanted to see in a future release. To have prompts to make increments in their goals was another area identified in which the app could be improved. In addition, if the cards were linked to trackers, they could work as reminders to drink more water or perform some more exercise.

Some would have liked to see the catch-up cards made more interactive, to increase the usability of the cards. In the survey data, 33% (16/48) of the participants could not test this functionality because they were not able to locate it within the app once these were missed as notifications. Users were not able to search for them again if the cards for that day had already been opened and seen once.

Many participants (16/48, 33%) could not understand the timing pattern of the card—at what time of the day cards appeared—and were unable to find the cards later in the app. Notifications were designed to appear 1 hour before the end of a scheduled shift and at 5 PM on a rest day. Currently, the cards are populated to enable appropriate notifications for the shift. For example, during a night shift, it might be focused on sleep, hydration, and relaxation. On a rest day, it might be focused on financial savings and family activity ideas.

**Discussion**

**Principal Findings**

**Overview**

As our previous study has highlighted, there are currently a number of health and fitness solutions available on multiple platforms in use by police officers and staff [15]. Notably, most apps focus on a particular element of health and well-being. They have not been designed to address the specific issues that police officers and staff face, nor are they configurable for the types of routines and working patterns that the users regularly encounter [31]. Recent literature reviews—such as a 2020 review of studies of gamification and mobile health (mHealth) apps for emergency service personnel (ESP) and police officers across 6 major databases—have highlighted a lack of literature in this area for these groups [32].
To meet the study objectives, where possible, user suggestions were implemented for improvements during the development process. The current app is designed to focus on the individual domains that comprise overall wellness—physical, mental, and financial well-being—of police personnel. Consideration of the shifts undertaken was at its center, as shift work patterns make tracking more difficult for this profession. Overall, feedback about the app was positive based on the SUS score received. Users particularly liked the shift manager, mood diary, trackers to log fluid intake and alcohol consumption, and the relevant supporting information made available for easy access. Approximately half (24/48, 50%) of the users thought that the app was easy to use, and this was considered to be the main reason for liking the app. However, there are still various improvements that need to be actioned. Some of these are improvements in shift manager, giving the ability to create bespoke shift patterns, functioning of the calendar feature of the existing trackers, and display of information to make it easy on the eye. Moreover, there are additional requests by participants to add new features to the app such as trackers to record activity and food intake and reminders to prompt them regarding fluid intake, sleep, and taking rest breaks.

**Food and Diet**

Police workers have great risk of being overweight or obese and risk of developing long-term health conditions [4]. Some of this risk is attributed to poor-quality diet (high in fat, sugar, and salt)—owing to the demands of shift work and the additional occupational stressors associated with police work modifying their relationship with food and unhealthy diet [33]. Police workers might not be able to plan where they might be at any point in a working day owing to the nature of the job, leading to disparate meals, few or small breakfasts, late mealtimes, and increased caloric intake at night [34]. This might also encourage them to make lifestyle choices based on whatever is easiest—for example, getting fast food on the go—rather than preparing healthy food in advance. Interestingly, a recent study by Kosmadopolous et al [34] observed that police officers had great intake of energy from fat and saturated fat during rest days and morning shifts than during evening or night shifts. However, the overall proportions of dietary macronutrients (fat, carbohydrate, and protein) did not significantly differ each day. Kosmadopolous et al [34] observed a series of dietary patterns that implicated the time at which food was consumed, rather than quantity or composition as the differentiating nutritional factor, which might affect metabolic health during shifts.

Working in shifts makes it difficult for police officers and staff to successfully adhere to and sustain healthy lifestyles in the long term. Participants suggested enhanced functionality regarding the ability to view a summary of how changing shifts could have an impact on the consumption of healthy food, and they were also in favor of using only 1 app to monitor all aspects of their health rather than many different ones. There are a number of existing dietary, nutritional, and food information apps available—such as MyFitnessPal [35]—but none of them successfully align completely with the lifestyle, fluctuating shift patterns, and demands of police work [1,2].

**Activity**

NHS guidelines in the United Kingdom [36] recommend that people should be performing some type of physical activity every day. Recent studies in this area involving 2 UK police forces have included a physical activity study using wearables, which used a combination of a Fitbit activity monitor and the “Bupa Boost” smartphone app to promote physical activity and reduce sedentary behavior in police officers [37,38]. Specifically targeted apps—focusing on cycling, running, or a combination of both—are not easily configurable for other activities such as swimming, owing to their design architecture. Buckingham et al [37] noted that there were large individual differences in preferences and perceived impact of the individual and social components of their intervention. These appeared to be owing to personal preferences and personality differences, rather than being associated with any identifiable characteristics, but they highlight how important personalization and tailoring are when considering activity and sedentary behaviors. The study [37] also emphasized that the targeted user group had accepted mHealth technology and found it extremely useful in improving physical health.

Notably, other existing technologies available for use, such as Police Fitness [39], which prepares individuals to pass the initial job entrance fitness exam or their annual fitness check, only concentrate on a particular aspect of fitness. Moreover, they do not naturally integrate with fluctuating shift patterns. The ability to track any kind of physical activity and to see a summary of the impact of shifts on activity levels was an aspect that participants would particularly like to see in a future iteration. Catch-up cards reminding them to participate in occasional physical activity (according to the scheduled shift) were something they considered would make a difference in their wellness journey.

**Sleep**

Among a group of US police officers, it was noted by researchers [40] that sleep disorders were common and significantly associated with increased risk of self-reported adverse health, performance, and safety outcomes. In a more recent study by Fekedulegn et al [41], which examined the association of shift work with sleep quality in police officers, the overall prevalence of poor sleep quality was 54%; 44% for the day shift, 60% for the afternoon shift, and 69% for the night shift. The study concluded that night and evening work schedules were associated with elevated prevalence of poor sleep quality among police officers.

For participants who worked shifts, maintaining regular sleep patterns was not always possible. Owing to the varied times they needed to go to bed and wake up, participants wanted an app that did not rely on standard workdays and the assumption that they would not be working on weekends and, importantly, also allowed them to integrate other features, such as similarly varying mealtimes when on shift and prompting reminders for relaxation before rest periods [15]. After testing the app, the participants expressed a desire to see a summary screen to help with understanding the impact of their shifts upon their sleep in helping to maintain overall wellness. For example, by logging
sleep patterns and then comparing and contrasting the types of mood after particular shift patterns.

Fluid Intake
Police officers and staff in frontline roles can sometimes find it difficult to maintain hydration status when on duty. Consequently, the effects of dehydration—fatigue, headaches, and irritability—can lead to loss of productivity while working [42]. This has led to campaigns to raise awareness such as the Take-a-Sip campaign—launched with funding from Police Care UK [42]. The NHS guidelines in the United Kingdom [43] recommend drinking between 6 and 8 glasses of fluid per day. Water, low-fat milk, and sugar-free drinks, including tea and coffee, all count toward this total, highlighting that intake does not necessarily need to be solely focused on water intake, which some existing apps might choose to focus on.

Participants mentioned that the goal-setting feature was useful when using it within the fluid intake section. There were suggestions to include this feature in other domains. In the long term, it is intended that, as recommended in the initial study [15], this feature, once finalized programmatically, is replicated and incorporated into other domains.

Good Mental Health
In terms of mHealth apps already targeting the mental well-being of police organizations, a UK-based mobile app—Backup Buddy—allows police officers in participating forces to informally view static audio and visual information and signposted support options about common mental health issues [44]. At the end of 2020, Thrive—a mental health and well-being app—was made available to 3500 officers and staff across West Staffordshire, with it also being made available to friends and family if required [45]. In addition, in the United Kingdom, the College of Policing recently conducted a randomized controlled trial, giving 1337 police officers in five forces access to either Headspace (a mobile mindfulness app) or Mindfit Cop (a web-based mindfulness resource). This study found that both resources improved well-being, life satisfaction, resilience, and performance compared with the control group. The authors concluded that the trial was sufficiently robust to provide evidence of well-being benefits [30]. A research study is also being conducted by the Police Federation of England and Wales to better understand police officer experiences of using the 87% mental well-being app. This app is designed to support employee well-being strategies [46].

By adding the good mental health section to the designs, a mood diary and the sleep tracker were made available to users to track personalized elements of mood. Integration of the mood diary with other sections of the app—such as hours of sleep recorded, exercise performed, skipped or eaten meals, and fluid intake—cross-referenced to individual shift patterns in the future would help to inform and provide users with a chance for reflection and therefore influence behavior change. In the long term, this component also has the potential to align more closely with national support services such as the National Police Wellbeing Service [47,48].

However, positives must be viewed alongside concerns from participants that were highlighted in this phase of testing regarding who would be able to access and make assumptions about the data they added. During testing, we were able to confirm that the data were confidential and that no one had access to personal information. This clarification was made to users to encourage use and allay participant concerns regarding how their managers might monitor and subsequently view some of the data entered. It has been noted that individuals in this profession feel reluctant to ask for help for themselves or to discuss their mental health with others. Historically, the police subculture has consisted of values involving masculinity, independence, and emotional control [20]. Such values may make it difficult for many police officers and staff to express emotion or seek mental health treatment, which places them at a disadvantage because internalizing their feelings might reflect in work performance [20]. In such a scenario, it becomes more critical for people in this profession to identify stress early, before it causes additional damage to their mental and physical health. The situation has been exacerbated during and after COVID-19, creating additional uncertainty and increasing the likelihood of stressful situations occurring [13,14].

Financial Well-Being
A 2006 study of Taiwanese police officers to assess the quality of life and prevalence of depression in police officers grouped together economic stressors including loans for a house or car, insufficient family income, and debt. According to their results, 52.2% (405/776) of male officers had economic stressors, whereas 30% (17/56) of female officers had economic stressors [21].

Compared with other professions, police officers and staff have different retirement policies; this means that they retire at a younger age than civilians and have more chance of experiencing negative impacts as they move toward retirement [22]. An Italian study found that retirees who were financially well-off were less likely to experience declining health when compared with those who were not [23]. Participants particularly welcomed the inclusion of a financial domain alongside the other aspects of the app, as it gave them good advice and ideas about the pension scheme offered, approaches to saving, and budget planning.

Shift Manager
Regarding activity, perceived pressure of work and organizational culture appear to be sturdy barriers to reducing sedentary time [49]. Previous studies have highlighted that there was a sense that high workload had resulted in working through breaks and during personal time becoming the norm [50]. Police staff has a mandated lawful requirement to take a break in their shift for which they are not paid, whereas police officers are paid for the full shift. This can mean that when operationally necessary, they work through breaks. Police officers in previous studies have expressed a need for more opportunities to take breaks and encouragement from managers or supervisors [37], as some of our previous participants had also noted [15].

Overall, 25% (12/48) of the users reported challenges in initially using shift manager. Some responded that they would continue to use their current apps (eg, Google) for managing their shifts, because of the additional functionality offered by other apps.
The ability to add personal notes to the shift manager and to share their weekly or monthly shift pattern with family members was something that users would like to see in an improved iteration. Having these features would cover users who preferred to use something different at the moment.

This feature has become the key central cog that other elements can be integrated with—for example, working in conjunction with the catch-up card function—therefore, the design and use of this function must be simple, effective, and easy to visualize and track. Feedback resulted in the early diary design to input shift being modified and renamed. Currently, there are 6 core shift patterns that have been prepopulated in the app, which users can choose to select. The frequency of catch-up cards was also set according to the shift chosen and included upon client feedback. The ability to view the effect of different factors (food, activity, fluid intake, and sleep) on overall wellness—in conjunction with shift—will prompt police staff and officers to make lifestyle modifications to improve health.

Feedback received from users also provided insights to the developers to improve the shift manager feature and to add a help section to educate beginners about how to add and edit their shift patterns.

**Goal Setting and Catch-Up Cards**

Behavior change is likely to play a large role in making an app focusing on health and well-being successful, with the suggestion that increased implementation of behavior change techniques could improve interventions and achieve high levels of user engagement [51]. The goal-setting feature was appreciated by participants when using it within the fluid intake section, and the intention is, once finalized, to integrate this functionality within other domains of the app.

Occupational stress is the main contributor to the risk of police officers and staff developing obesity and increasing the risk of long-term health conditions such as cardiovascular and metabolic diseases. To control this, it is very important to self-manage sleep patterns, take regular breaks, monitor dietary and food habits, regulate water intake, help with smoking cessation, and reduce alcohol consumption. Referring to the study conducted by Voyer [52], nudging offers a choice pattern to users that can make changes in their behavior and how decisions are taken at a personal level to improve health. Therefore, it is considered as a behavioral economic concept in the new world. The study by Kwan et al [53] highlighted that reminding the patient to maintain good habits by sending nudges has the potential to reduce the cost of health care and help patients take long-term control of their health. Self-management empowers users to make decisions in favor of their health, and nudges or reminders keep them more involved and informed in maintaining their health. Nudge theory is a young behavioral economic concept that “influences the behaviour and decision-making of patients through choice architecture” [52]. Nudging is not mandatory; rather, it gives small choices in behavior, at a level that has the potential to influence.

In the initial research requirement gathering, participants mentioned that notifications on the app would help them to perform tasks that they generally forgot, as they did not access the app many times during the day [15]. Catch-up cards were introduced and linked to the shift manager in this version of the app to support and encourage police officers and staff to try and maintain good habits. Catch-up cards are scheduled differently for each shift pattern and rest day to prepare the participants for their next shift. They work as a “nudge” or reminder for the participants to do breathing exercises, sleep on time, complete activities, and so on.

**Supplementary Information and Guidance**

Previously, some pilot research work was undertaken with project team members on an app to centralize evidence-based nutrition and lifestyle guidance for health care professionals and people living beyond cancer—enabling them to obtain guidance and information and create and track nutrition and activity-related goals [54]. The research work also fed useful reflections into the design of this project. Each domain in the app that requires input from the users has an information section that gives recommendations about how much fluid to take, what the healthy options are, and so on, according to NHS guidelines. Participants found this guiding information to be helpful, as it was easy to refer to in case of query. However, in this study, there were few suggestions received around improving the design to draw more attention to these particular sections.

**Limitations**

Although the authors believe that the results from this study can be generalized to other police forces across the country, we acknowledge that there is a limitation of only accessing data from a subsection of 1 regional UK police force, which might have inherent organizational biases toward health and well-being. However, this approach can be expanded to cover more regions in due course. Regarding bias with purposeful sampling—where the belief is that qualitative research should be describing the medium or the norm—the point to underline is that new phenomena are being described; therefore, we needed to purposively select the best examples of what we were interested in. This gave us the clearest cases with the least “noise” or extraneous errors and allowed for the identification of characteristics and boundaries [55]. A further limitation of the study was the greater number of responses from women (32/48, 67%) than those from men when compared with regional UK police workforce numbers as of March 2022, which noted that the gender split was 48% women and 52% men [56].

**Future Studies**

The wellness app for the commissioning force has now been subsequently soft launched with funding for the bespoke pattern design agreed upon. The findings from this study are being shared with the developers to analyze the possibilities of implementing them in the next release of the app and have the improved version of the app trialed before its final launch for a wide police population.

**Organizational Support**

In addition to a well-designed mobile app, suitable scaffolded organizational support implemented alongside the eventual release of the app, together with an evaluation of how the app is being used, offers the best chance of producing positive results. This could include encouraging the use of more

https://humanfactors.jmir.org/2023/1/e42912 JMIR Hum Factors 2023 | vol. 10 | e42912 | p.159 (page number not for citation purposes)
health-oriented leadership—a style of leadership associated with more well-being and low levels of burnout, depression, and physical complaints among police officers [57].

**Voice Integration**

Emergent technological enhancements (such as Google Voice and Google Assistant) offer opportunities for improved personalized eHealth solutions [58] and increased engagement [59] and adherence [60]. In the area of nutrition, projects are already being conducted, such as the design, development, and evaluation of an Alexa Skill on food and nutrition management for native American patients with diabetes [61]. Similarly, regarding fitness, TandemTrack combines a mobile app and an Alexa skill to support exercise regimens, data capture, feedback, and reminders [62]. Members of this team have previously explored the tentative use of voice-activated speakers or assistants such as Google’s Assistant or Amazon’s Alexa to enable the input of information via voice rather than keyboard [63], which might also assist in making this type of innovation quick to use and therefore more user-friendly.

**Investigating Other ESP Pathways**

These unique issues, the continuing long-term effects of the pandemic on employee health and well-being, and the highlighted gaps in solutions available at present are not only applicable to police staff and officers but also to other ESP—such as firefighters [64,65], paramedics [66,67], and health care professionals [68,69]—and other shift workers. Therefore, the project team is seeking further funding, to explore the development and expansion of this approach to health and well-being issues in other sectors, in addition to the police.

**Conclusions**

This study highlights the importance of coparticipation with officers and staff across the entire development cycle, to coproduce a human-centered design methodology to enable the development of a considered and user-centered solution. It demonstrates the need for producing a multifunctional tool rather than focusing purely on an individual element for this user group. It also highlights how linking and being able to track optional, personalized elements of health data against each other, cross-referenced to individual shift patterns, might help to inform and provide users with a chance for reflection and therefore influence behavior change.

**Acknowledgments**

The authors would like to thank all the participants of the web-based survey and interviews. This research project was cofunded by Bournemouth University and the commissioning client.

**Authors’ Contributions**

JM, FB, and HD conceived and designed the study. RM conducted the participant interviews and data analysis. RM led the preparation of the paper. All authors critically reviewed and revised the paper and participated in its final approval.

**Conflicts of Interest**

None declared.

**References**


15. Swanston E, Pulman A, Dogan H, Murphy J, Bitters F. Scoping the need for a tailored mHealth app to improve health and well-being behavioral transformation in the police: exploring the views of UK police workers via web-based surveys and client meetings. JIMR Form Res 2021 Aug 31;5(8):e28075 [FREE Full text] [doi: 10.2196/28075] [Medline: 34463625]


44. Backup Buddy UK homepage. Backup Buddy UK. URL: https://backupbuddy.com/about [accessed 2023-06-29]


46. 87% app pilot study participant information. Police Federation. URL: https://www.polfed.org/87percentpilotstudytetms [accessed 2023-06-29]

47. National police wellbeing service. Oscar Kilo.: College of Policing Ltd URL: https://oscarkilo.org.uk/ [accessed 2023-06-29]


 Abbreviations

ESP: emergency service personnel
mHealth: mobile health
MoSCoW: Must Have, Should Have, Could Have, Won’t Have
NHS: National Health Service
SUS: System Usability Scale

Edited by A Kushniruk; submitted 04.10.22; peer-reviewed by I Shubina, A Hadianfard; comments to author 17.04.23; revised version received 24.04.23; accepted 30.04.23; published 04.08.23.

Please cite as:
Mehra R, Pulman A, Dogan H, Murphy J, Bitters F
A Tailored mHealth App for Improving Health and Well-Being Behavioral Transformation in UK Police Workers: Usability Testing via a Mixed Methods Study
JMIR Hum Factors 2023;10:e42912
URL: https://humanfactors.jmir.org/2023/1/e42912
doi:10.2196/42912
PMID:37540549

©Richa Mehra, Andy Pulman, Huseyin Dogan, Jane Murphy, Fiona Bitters. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 04.08.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Bridging the Communication Gap Between People With Cognitive Impairments and Their Caregivers Using mHealth Apps: User-Centered Design and Evaluation Study With People With 22q11 Deletion Syndrome

Martijn Van Dooren¹, MSc; Robin De Croon¹, PhD; Ann Swillen²,³, PhD; Katrien Verbert¹, PhD

¹Department of Computer Science, KU Leuven, Leuven, Belgium
²Department of Human Genetics, KU Leuven, Leuven, Belgium
³Center for Human Genetics, University Hospital Gasthuisberg Leuven, Leuven, Belgium

Corresponding Author:
Robin De Croon, PhD
Department of Computer Science
KU Leuven
Celestijnenlaan 200A
Leuven, 3001
Belgium
Phone: 32 16 37 39 76
Email: robin.decroon@kuleuven.be

Abstract

Background: In families with children with cognitive impairments, both parents and children experience tension and have questions because of a lack of communication and adequate information. Therefore, there is a great need to develop tools that can help bridge the communication gap between patients and caregivers by stimulating conversations and providing psychoeducational tools. mHealth apps show great potential in this context.

Objective: The objective of this research is to discover the specific ways young people with cognitive impairments and their families interact with mHealth apps in the context of bridging the communication gap. This newly discovered information leads to potentially more impactful mHealth interventions in the future. Therefore, this paper documents the design and development of a mHealth app for a specific group of people with cognitive impairments—people with 22q11 deletion syndrome (22q11 DS)—and their caregivers, as well as key learnings from the evaluation of this app.

Methods: An iterative, user-centered design approach is used to design and develop the app. Design and evaluation happens in 2 phases. During the design phase, feedback is gathered from 2 medical experts and 3 human computer interaction (HCI) experts using a low-fidelity paper prototype. During the evaluation phase, feedback is gathered from 8 families with a child with 22q11 DS using a fully working proof of concept. This phase consists of a semistructured interview, a 2–4–week trial period, and a concluding semistructured interview.

Results: The evaluation results of the fully working proof of concept led to design recommendations related to four different topics: (1) overcoming usage barriers, (2) stimulating conversation through a mHealth app, (3) providing information, and (4) bringing continual added value. Results are presented according to six different categories obtained in a thematic analysis: (1) feedback about the app “as is,” (2) difficulties, (3) comparison between physical and digital tool, (4) extensions, (5) intention, and (6) other.

Conclusions: In this research, the need for apps that help bridge the communication gap between a person with cognitive impairment and their caregiver is confirmed. All participating families express their gratitude and mention the added value for other families. Therefore, it is highly encouraged for clinics and institutions to take action and develop an app to be used in practice. Furthermore, considerations when developing for people with 22q11 DS, or more broadly, people with cognitive impairments, are proposed. First, one should keep design principles in mind to overcome usage barriers. Next, recognition is a key concept when stimulating conversations through mobile apps. Third, information should be provided by a trusted source, and more than just clinical information can be considered valuable. Finally, having the possibility of using a digital tool that can be personalized brings continual added value.
Introduction

The Need for Bridging the Gap

During the last few decades, awareness and recognition about different genetic syndromes and genomic disorders have significantly increased. Specifically, for both 22q11 deletion syndrome (22q11 DS) and autism spectrum disorder, this has led to increased prevalence rates [1-4]. For Down syndrome, prevalence rates have been rising as well due to an increase in average maternal age [5,6]. People with both 22q11 DS and Down syndrome or autism spectrum disorder all have a high probability of having social and cognitive impairments [7-10]. Therefore, the increase in attention, awareness, and prevalence of these syndromes and disorders has led to increased attention for the needs of people with social cognitive impairments.

One subgroup of people with cognitive impairments will be the target group of this paper: young people with 22q11 DS. 22q11 DS is a congenital syndrome caused by a deletion or duplication on the long arm of chromosome 22 [11]. The prevalence rate of 22q11 DS is about 1 in 4500 [12], which makes it a rarer genetic syndrome than Down syndrome, which has a prevalence rate of about 1 in 720 [6]. People with 22q11 DS often have several social cognitive impairments, of which impaired emotion processing, circumscribed interests, deficits in sharing attention, gestural communication, initiating and maintaining conversations, and poor adaptive socialization are some examples [13,14].

Besides this, research states that “individuals with 22q11 DS seem to be aware of their health and psychological problems, but on the questions about social relationships and environment, they (possibly) respond with socially desirable answers. Individuals with 22q11 DS often want to please other people and do their very best in any circumstances. It is possible that they don’t want to bother anyone with their difficulties in social relationships and interaction with their environment” [15]. Combining all this, it can be concluded that it is not easy to have meaningful conversations about feelings, experiences, and symptoms with children with 22q11 DS. The lack of adapted and adequate communication can lead to frustration between people with 22q11 DS and their caregivers.

Bridging the communication gap between people with cognitive impairments and their caregivers is therefore a challenge present in many families today. The rise of modern technology potentially holds solutions to this challenge. mHealth apps could potentially be a great tool for supporting communication in these situations.

mHealth Apps: a Viable Solution

In recent years, research has been done in the context of both mHealth apps for people with cognitive impairments and bridging the communication gap between people with cognitive impairments and their caregivers. First, literature indicates that families of patients with Down syndrome, Williams syndrome, and 22q11 DS showed a positive attitude toward mHealth technologies [16]. Besides, parents of children with 22q11 DS indicated they could have benefited from additional support to increase their confidence and success while disclosing the diagnosis to their child. Also, it could have increased the child’s comprehension of the information [17]. Combining these arguments, it might be valuable to investigate using mHealth apps with the specific goal of supporting communication for people with 22q11 DS and their caregivers.

Besides this observation, other mHealth apps that support families for other target groups were shown to have a high possibility of successful outcomes. In a family adaptation program for children with Down syndrome, all parents indicated they were willing to recommend this form of intervention to other families [18]. When using an mHealth resource for caregivers of cancer patients, these caregivers found the app highly useful in their experience of caregiving [19]. Lastly, in a scoping review to inform the development of mHealth apps for families with a child with Down syndrome, it was concluded that effective care coordination through such an app has the potential to increase family satisfaction [20].

Previous research also shows that developing mHealth resources specifically for people with 22q11 DS and other cognitive impairments has a great chance of helping them succeed in their goals. First, a remote cognitive remediation program with 22q11 DS youth was implemented without any problems [21]. This highlights the feasibility of any form of remote intervention for people with 22q11 DS. Furthermore, people with Down syndrome showed there are no hurdles to using any sort of touch gesture on a touchscreen [22,23]. As people with 22q11 DS likely have the same or better motoric abilities, this is an essential argument for the viability of an mHealth resource to support people with 22q11 DS.

Next, design implications were proposed to increase the potential success of mHealth apps for people with cognitive impairments. While designing these kinds of apps, keeping it simple, using visual cues, avoiding complex login functionality, using personalization, keeping patients’ mental models in mind, and employing a dynamic difficulty level are essential things to consider [24-26].

Combining all this, little research has been done on developing mobile mHealth apps for people with cognitive impairments and their caregivers in the context of bridging the communication gap between these 2. When considering people with 22q11 DS specifically, the need for research into these topics is even greater, as almost no research has been done concerning these matters. However, as all necessary building blocks are readily available, this study will focus on developing an mHealth app for people with 22q11 DS and their caregivers to help bridge the communication gap.

KEYWORDS

22q11 deletion syndrome; 22q11 DS; cognitive impairments; communication gap; mHealth
Study Objective

The objective of this study is to gain further insight into how to build successful apps that support people with cognitive impairments and their families in the experience of caregiving. This study will focus specifically on people with 22q11 DS. The following research questions arise:

• Question 1: how can an mHealth app lower the communication burden between people with 22q11 DS and their caregivers (family and close friends)?
• Question 2: how can an mHealth app be a stimulant for people with 22q11 DS and their caregivers to have more regular conversations about the syndrome?
• Question 3: how can clinical information about the syndrome and the clinical symptoms of the condition be presented to young people with 22q11 DS and their caregivers to enhance health literacy?
• Question 4: which are the most important design principles when developing an app for young people with 22q11 DS?

A fully working proof of concept is designed, developed, and evaluated with young people with 22q11 DS and their families to formulate an answer to these research questions. The work in this paper contains valuable contributions in 2 areas. First, important design principles when designing for people with cognitive impairments, more specifically people with 22q11 DS, contribute to the health care informatics domain when considering mHealth apps for people with cognitive impairments. Besides this, important contributions are made to the psychoeducational domain by providing further insight into how to maximize the potential of a digital tool like the one created in this research.

Methods

In the next parts, the full methodology used in this research is explained, referring to the overall study design, the participants in the research, the way data is analyzed, and the ethics approval granted.

Study Design

This research incorporates an iterative, user-centered design process. By dividing the design process into different phases, insightful feedback from both experts and users is gathered. This research is split up into 2 main phases: a design phase that incorporates a low-level prototype and an evaluation phase that incorporates a fully working proof of concept. The latter again consists of 3 different parts: an initial interview that incorporates a first version of the proof of concept, a trial period, and a concluding interview. A visual overview of this study design can be found in Figure 1.

As a starting point for this research, a physical tool that was developed to support families with children with rare genetic syndromes is used as a starting point. The physical tool “Together we put the puzzle” launched in March 2020 [27]. The tool is currently regularly used in genetic counseling, mainly by clinical orthopedists, when parents and children need psychoeducation about their syndrome. Besides this, 50 families use the tool at home, and about 30 early intervention services, clinical genetic centers, special education schools, and rehabilitation centers are currently working with the psychoeducational tool (puzzle and booklet). Since April 2022, the tool has also been available in English, and 50 different copies have been sent to medical doctors (clinical geneticists, neuropediatricians, psychiatrists, etc) and allied health professionals (psychologists, remedial therapists, etc) working in the field of neurodevelopmental disorders due to a copy number variant (NDD-CNVs) across Europe.
Figure 1. Visual overview of the study design.

**Design Phase: evaluation with HCI experts and medical experts**

- Interviews incorporating low-fidelity paper prototype

**Evaluation phase: evaluation with young people with 22q11 DS and their family**

- Interviews incorporating a proof of concept
- Trial period
- Concluding interviews

**Design Phase: Evaluation With Medical Experts and Human-Computer Interaction Experts Using a Low-Fidelity Paper Prototype**

Using the concepts upon which the physical tool “Together we put the puzzle” is based, 2 prototypes are developed on paper. These low-fidelity prototypes are evaluated with both medical experts who have experience in the treatment of children with 22q11 DS and use “Together we put the puzzle” in practice and with experts in the human-computer interaction (HCI) domain. The latter was done to gain insight into the most prominent usability issues. The feedback was gathered in a web-based one-to-one think-aloud session of 45 minutes in which the 2 prototypes were shown to the users.

**Evaluation Phase: User Evaluation With a Fully Working Proof of Concept**

The second phase consists of 3 different parts. By conducting both interviews and allowing for a trial period, both qualitative and quantitative results are collected and analyzed. All different parts are conducted with the same participating families. The next paragraphs elaborate further upon the different parts of this evaluation phase.

**Interviews Incorporating a Fully Working Proof of Concept**

Based upon the evaluation of the paper prototype, a fully working proof of concept is developed using Meteor.js as the underlying cross-platform web architecture. This ensured the proof of concept was compatible with a variety of devices, such as smartphones and tablets, and supported all popular operating systems (eg, iOS and Android). Moreover, the proof-of-concept supported offline caching to prevent any network connectivity issues. Individual semistructured interviews of 60 minutes with young people with 22q11 DS and their parents are conducted at their own homes.

**Trial Period**

After the initial first interviews, the families can use the proof of concept in a trial period lasting 2-4 weeks until the next interview. Families are asked to use the app at least once during this period.

**Concluding Interviews**

During a second individual, semistructured interview session of 45 minutes, families give final feedback. These interviews again take place at their own houses. In these concluding sessions, new insights can be gained after considering the possibility of using the app during a trial period, and feedback from the earlier interview sessions might be confirmed further.

**Participants**

During the 2 phases of the study, different groups of participants take part in the study. In the design phase, both medical experts in the field of 22q11 DS and diseases that lead to cognitive impairments and HCI experts are involved. In the evaluation phase, young people with 22q11 DS and their parents are involved.

The medical experts that take part in this research are 2 medical experts that have a proven track record in the field of 22q11 DS and diseases with other cognitive impairments. Besides this, the feedback of 1 female and 2 male HCI experts is gathered in the design phase.

The group of people participating in the evaluation phase are young people with 22q11 DS and their parents and siblings. These young people are required to be between 8 and 23 years of age and need to have taken part in a physical session where the original puzzle was used at least once in their previous treatment. The latter is important, as this research does not want to focus on the contents and workings of the resource and does
not want to intervene medically. The parents of the young people did not need to adhere to specific conditions.

**Data Synthesis and Analysis**

During the design phase, feedback is gathered about the workings of the low-level paper prototype. Difficulties, possible extensions, and positive feedback are part of this feedback. These findings, in combination with the low-level paper prototype, formed the basis for the development of the proof of concept.

A total of 2 main types of results are acquired in the evaluation phase of the research. First, qualitative results are obtained from both interviews. Besides this, additional quantitative data is obtained from the logs that are collected during the use of the app in the trial period.

Qualitative results are analyzed using thematic analysis [28]. Results are collected and categorized according to the following themes: (1) feedback about the app “as is,” (2) difficulties, (3) comparison between physical and digital tool, (4) extensions, (5) intention, and (6) other.

Quantitative analysis was used to answer questions about the average session length of users and all the different functionalities that were or were not used by families during the trial period.

**Ethics Approval**

As this study involves vulnerable participants due to their medical condition and cognitive impairment, ethics approval had to be given by the Ethical Committee for Research at KU Leuven and UZ Leuven to conduct this research. The committee approved the study in March 2022, and it is identified by S-number S66151. Besides approval by the committee, informed consent was obtained from all participants.

**Results**

**Overview**

This section discusses the most important results from the different parts of the evaluation phase of the research. For the evaluation phase, a total of 8 families were recruited. Table 1 presents more specific details about the different families.

<table>
<thead>
<tr>
<th>ID</th>
<th>Age of the child (years)</th>
<th>Gender of the child</th>
<th>Brothers and sisters, gender, and age (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9</td>
<td>F&lt;sup&gt;a&lt;/sup&gt;</td>
<td>F 6</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>M&lt;sup&gt;b&lt;/sup&gt;</td>
<td>M 1, M 3, M 5, F 10</td>
</tr>
<tr>
<td>3</td>
<td>18</td>
<td>F</td>
<td>M 13, F 16</td>
</tr>
<tr>
<td>4</td>
<td>22</td>
<td>M</td>
<td>F 22</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>F</td>
<td>F 7</td>
</tr>
<tr>
<td>6</td>
<td>18</td>
<td>F</td>
<td>F 13</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>F</td>
<td>F 7, M 11, M 13</td>
</tr>
<tr>
<td>8</td>
<td>16</td>
<td>M</td>
<td>F 13</td>
</tr>
</tbody>
</table>

<sup>a</sup>F: Female.

<sup>b</sup>M: Male.

**Interviews Incorporating Proof of Concept**

**Overview**

During the first semistructured interviews, the first version of the fully working proof of concept is used. An overview of this proof of concept can be found in Figure 2. The app exists out of an onboarding process, functionalities for parents and siblings, functionalities for the person with 22q11 DS, and the possibility to collectively put the puzzle. The functionalities for children exist out of getting answers to frequently asked questions, rating themselves on some skills with stars, and personalizing the app by choosing a color and an avatar. The feedback on the proof of concept is presented according to the categories mentioned earlier.
**Figure 2.** Some of the most important screens in the app: (A) the home screen with links to the 3 main parts, (B) the different functionalities for a child, (C) the way answers to questions are presented to children, (D) the puzzle overview screen, (E) the list of possible puzzle pieces, (F) the way a puzzle piece is presented to a child, (G) the option to make a puzzle piece larger or smaller depending on the level of recognition, and (H) selection screen for parents, brothers, and sisters to navigate to their own part of the app.

**Feedback About the App “as is”**

Using the app in general is easy for every child that participates in the study (8/8, 100%). The possibility to choose your own color and avatar is highly appreciated by 87.5% (7/8). Children focus heavily on the visible part of the screen; when presented lists to scroll through, children often choose 1 of the visible parts and seem to minimize scrolling (5/8, 62.5%). If buttons are not visible on the screen immediately, some confusion arises in a few cases (2/8, 25%). Besides this, audiovisual resources show a high impact and get the preference of all the children but one (7/8, 87.5%). Overall, the app receives highly positive feedback from both children and parents, who acknowledge the value such an app can have (8/8, 100%). A striking example of this is one of the younger children (aged 10 years) answering the final question, “Do you have any additional comments you want to add to our conversation?” with, “Will you not forget to let us know where we can find the app so I can use it in the..."
future?" or multiple parents mentioning, “I’m sure this will help a lot of other families.”

**Difficulties**

The main difficulties that arise can be classified into 2 main themes. First, everything text-related should be thoroughly thought about. Difficult words and long sentences cause problems and a loss of engagement during the use of the app, mainly for younger children (5/8, 62.5%). Besides this, as mentioned before, when buttons are not immediately visible on the screen, this can cause confusion as well (2/8, 25%).

**Comparison Between Physical and Digital Tool**

When comparing the physical and digital tool, families sometimes explicitly mention preferring the digital version (4/8, 50%). The burden of using this tool is lower, often because of practical arguments. Using a physical tool simply requires more effort and energy. One of the parents states:

*Just having to walk to the closet in the other room and taking the puzzle out is already a burden to use it whereas this is not the case with a mobile application.*

Besides, families indicate that the fact that the puzzle is saved creates new opportunities. As everything that is done in the app can be undone, families also indicate they would use functionalities quicker than in the physical case (3/8, 37.5%). For example, only 4 blank puzzle pieces are provided in the physical tool, whereas these are unlimited in the app. This leads to families being less afraid to create their own puzzle pieces. Finally, some of the families indicate they do not think a phone is the appropriate medium to use as a tool for communication support within the whole family because of the simple reason that the screen is too small. Being able to use the app on a tablet or even on a computer could solve this issue (3/8, 37.5%).

**Extensions**

The main extensions that come up are additional information for parents, brothers, and sisters (8/8, 100%), the possibility for brothers and sisters to lay their own puzzle (3/8, 37.5%), and introducing a feedback system for asked questions in the app (3/8, 37.5%).

**Intention**

All children but 1 indicated they were interested in further use of the app during the interview itself (7/8, 87.5%). Parents also say they see the additional benefits. In families where children are already older, they indicate the need for the app is not as high, but they do see the value in a similar app for families with younger children (1/8, 12.5%).

**Other**

During the interviews, the Facebook group of the parent association is mentioned multiple times. However, the subjective and more negatively focused nature of this information leads to a lot of people not wanting to be active in this context. They mention the fact that they would have more trust in an app created by a trusted third party like a hospital or government institution (4/8, 50%). Finally, in every family, at the end of the interview, parents emphasized the importance or added value of this kind of research (8/8, 100%).

**Quantitative Results From the Logs Generated During the Trial Period**

While evaluating the usage of the app by looking at the generated logs, a few things became very apparent. A total of 2 families with older children (both 18 years of age) did not use the app during the trial period, even though this was asked at the end of the previous interview and in an email that was sent shortly after the interview. However, families that did use the app, half of the sessions were 26 minutes or longer, with 2 sessions even lasting 53 and 61 minutes. No technological issues were reported or logged. An overview of the different sessions per family and their duration can be found in Figure 3.

When looking at the functionalities that were used by children and their parents, it can be observed that, overall, the puzzle got a lot of attention.

**Figure 3. Sessions per family.**
Concluding Interview

The families with older children that did not use the app during the trial period indicated they did not have the need to do so. As the syndrome is not an active subject anymore within the family, the need for communication support apps is also lessened. However, they both indicated that if a more urgent situation came up, they would use the app as a tool to support them in their conversations.

In families that did use the app, it was spontaneously mentioned that they had talked about things they had never talked about before. However, one family indicated that right now the puzzle was not yet interactive enough to keep the children’s attention while using it. No additional difficulties came up, except for the fact that one time the child did not understand why the puzzle piece could be made larger or smaller. Children keep preferring the digital puzzle. Finally, one child proposes to extend the app with the possibility of being able to capture pictures themselves to use as images for the puzzle pieces.

Last but not least, it was further confirmed that there is a great need to further involve siblings. Giving them an equally important role in the app is a step forward. There is also a great need for informing siblings, and besides, the ability for every member to put the puzzle together based on their own experiences creates starting points for new, valuable conversations.

Discussion

Principal Findings

To summarize our findings, young people with 22q11 DS and their families highly value the developed mHealth app as a supporting tool in communication and for gaining additional information. This confirms the need for these kinds of solutions for families with children with cognitive impairments [15,29-31]. It is highly encouraged that institutions like hospitals or governments take action and start the development of this kind of tool. One should keep in mind that combining powers is a better approach than developing stand-alone apps that all need individual maintenance. This is especially true for apps such as the one discussed in this research, as it shows value for a lot of different target groups.

In general, people with 22q11 DS show few difficulties in using the developed mobile app. This confirms the conclusion from other research that remote interventions with this target group can be successful [21]. It also confirms the earlier presumption that no motorical difficulties would arise, as children with Down syndrome were shown to be able to use mobile apps as well in earlier research [22,23].

Furthermore, while doing interviews with people with 22q11 DS, a couple of things stood out. The fact they are pleasers [15] shows when answering the questions in the interviews. It happens regularly when a child indicates they understand something, whereas if asked to perform a certain action, it becomes clear they do not understand this at all. The verbal IQ of these children is often higher than their performance IQ, but this holds the potential risk of overestimating their capacities [32,33].

One of the strengths of this research is that it combines the confirmation of an important need with concrete considerations when developing for people with cognitive disabilities, like people with 22q11 DS. The latter are extensively discussed in the following paragraphs.

Considerations When Developing Mobile Apps for People With Cognitive Impairments

Overview

This research provides insights that should be considered when developing a mobile app for people with cognitive disabilities in the context of bridging the communication gap. These insights can serve as general guidelines. The insights are discussed according to the three-step structure displayed in Figure 4 below: (1) what are the most prominent usage barriers to using mHealth apps, and how can one overcome these? (2) How to achieve desired results, in this case, lowering the communication burden within families and providing information to families? (3) What are the important implications in the context of having continual added value for the target group?
Figure 4. Schematic overview of considerations when developing for people with cognitive impairments in the context of bridging the communication gap.

**Usage Barriers**

First and foremost, from this research, it is clear that if the guidelines for developing for the web for people with cognitive impairments are followed when developing mobile apps [34,35], usage barriers are limited. However, one area where it is necessary to be extra cautious is everything related to text. The research shows that using too much text or too difficult words leads to a loss of engagement and interest among young people with 22q11 DS. However, other important design considerations are shown to be important to make sure the mobile app can be easily used. When designing specifically for young people with 22q11 DS, these considerations should be taken into account:

- Limit the actions needed to 1 action per screen.
- Keep all the necessary information directly visible on the screen without scrolling.
- Use audiovisual means wherever possible.
- Avoid long sentences and large collections of text.
- Pay close attention to the words you use; the easier, the better.
- Use grids over lists (earlier research concluded this as well [36]).

Finally, specifically in the context of lowering the communication burden by providing a tool that can be used together with the whole family, some participants indicate that a mobile phone might not be the right medium due to the limited screen size. Families prefer to use a larger screen in this specific situation; for example, using a tablet offers more potential in this area.

**Achieving Desired Results: Lowering the Communication Burden**

With the goal of lowering the communication burden, the concept of recognition played a key role. By creating points of recognition using a mobile app, conversation starters are offered to families to talk about more difficult subjects. It was indicated by participants that having these starting points for conversations is in itself enough to lower the communication burden. To maximize the lowering of the communication burden, one can look for various ways to introduce these points of recognition and conversation starters, not only through the existing puzzle pieces, but also, for example, by including testimonials and videos of other people with the same syndrome.

The concept of recognition that appears can be found in tools created for other target groups as well. One could argue that reminiscence is a specific kind of recognition. For example, stimulating reminiscence through technology with older adults is found to have a positive effect on communication both in people with and without cognitive impairments [37,38]. Similarly, in this research, using technology to make people think not about situations in the past but about situations in the present appears to be an important element that can stimulate conversations.

Besides, in this research, the very important role played by the brothers and sisters of the child with 22q11 DS appeared. It is important to involve these siblings heavily, as they both have their own questions and challenges but are also some of the people who know the child with 22q11 DS best [39-41]. This research implements some recommendations for practice, like encouraging siblings’ curiosity about the mindset of their brother or sister with a disorder and inviting the sibling to discuss issues regarding feeling normal and feeling different [41]. It also confirms the fact that feedback from siblings is highly valuable during research itself [41].

Finally, in the field of communication, it should be noted that families tend to use a mobile app on a more individual basis than a physical tool. Conversations occur not only synchronously while using the app together but also asynchronously. For example, children find things they think are interesting while...
using the app and afterwards tell their parents about them and start new conversations.

**Achieving Desired Results: Providing Information**

For the goal of providing information to people with 22q11 DS and their caregivers, one should be conscious of the fact that not all children with 22q11 DS have the need to have a deeper understanding of why things are the way they are. This is largely attributed to having cognitive impairments [42,43]. However, when presenting information to them, using video and information at their level of thinking are crucial aspects to being successful in this goal.

When presenting clinical information to parents, a question-answer system split up into different categories seems like a potential way to go. What they think is especially important is a trusted third party that provides the information. Therefore, it should be encouraged that official institutions with knowledge about 22q11 DS are the creators of these kinds of apps and provide the necessary maintenance. An important consideration for practice also involves the fact that parents are not only interested in clinical information. They are as interested in the practical consequences of having to manage a child with 22q11 DS regarding taxes, institutions, support organizations, and other related topics.

**Bringing Continual Added Value**

Having continual added value has been shown to be a complex topic in this research. From the feedback and effective usage of the app, it becomes clear that not every family needs regular conversations about the syndrome. What families need are tools that can support them at the moments they need them. In families with younger children, this will be a more permanent situation, whereas in families with older children, only at the most urgent points in time will an app be used.

However, just having the option of using a digital tool like the one in this research is in itself already a way to bring continual added value. Families indicate they would more quickly use a digital tool than a physical tool, solely because of the lower practical burdens of using it.

Last but not least, in order to boost engagement, personalization has been shown both in previous research [44] and in this study to be an important aspect for the success of an app.

**Limitations**

A few important limitations need to be pointed out in this research. First, although 8 different families with children with 22q11 DS took part in the research, the quantitative analysis of the logs with more participants could lead to even more valuable insights. Nonetheless, our participants provide important perspectives that enhance our understanding of their situation. Involving the families provided additional information and firsthand knowledge gained through years of experience and interaction with health professionals. We employed a mixed methods approach, combining qualitative methods like in-depth interviews with quantitative data after real-world usage, enabling a comprehensive exploration of participants’ experiences and needs. However, the trial period in this study can be perceived as relatively short. On the other hand, the trial period did provide us with valuable additional insights that would not have been collected with interviews alone. This study serves as an exploratory investigation, laying the groundwork for future research and informing studies with larger sample sizes, leading to a gradual expansion of knowledge in the field.

**Conclusions**

In this research, an iterative, user-centered design process is carried out. Based on the design phase, a fully working proof of concept of a mobile app is developed with the goal of bridging the communication gap between people with 22q11 DS and their caregivers. This proof of concept is evaluated during the evaluation phase. The need for these kinds of apps is confirmed. All participating families express their gratitude and mention the added value for other families. Therefore, it is highly encouraged for institutions to act and develop an app to be used in practice. Furthermore, considerations when developing for people with 22q11 DS, or more broadly, people with cognitive impairments, are proposed. First, one should keep design principles in mind to overcome usage barriers. Next, recognition is a key concept when stimulating conversations through mobile apps. Third, information should be provided by a trusted source, and not only clinical information brings added value. Finally, having the possibility of using a digital tool that can be personalized brings continual added value.

**Acknowledgments**

The authors would like to thank all the patients and their families for their participation in the project.

**Conflicts of Interest**

None declared.

**References**


44. Inne S. Het gebruik van een planningsapplicatie bij jongeren met het 22q11 deletie syndroom. Dissertation Master in de
41. Fjermestad KW, Haukeland YB, Mossige S, Vatne TM. Children’s perspectives on the experiences of their siblings with
40. Ross P, Cuskelly M. Adjustment, sibling problems and coping strategies of brothers and sisters of children with autistic
38. Rose N, Whitworth A, Smart S, Oliver E, Cartwright J. “I remember when …”: the impact of reminiscence therapy on
36. Wahed SA, Adam H, Kuch J, Palzer V, Schnalzer B, Hanke S. QUALY: a money management application for people with
35. Making content usable for people with cognitive and learning disabilities. World Wide Web Consortium (W3C). URL:
34. WCAG 2 overview. W3C Web Accessibility Initiative (WAI). URL: https://www
32. Antshel KM, Fremont W, Kates WR. The neurocognitive phenotype in velo-cardio-facial syndrome: a developmental
31. Kuenzel E, Seguin D, Nicolson R, Duerden EG. Early adversity and positive parenting: association with cognitive outcomes
29. Ben-Sasson A, Soto TW, Martínez-Pedraza F, Carter AS. Early sensory over-responsivity in toddlers with autism spectrum
23. WAAG 2021. URL: https://kuleuven.eje."\end{verbatim}

Abbreviations

22q11 DS: 22q11 deletion syndrome
HCl: human computer interaction
NDD-CNv: neurodevelopmental disorder due to a copy number variant

Edited by A Kushniruk; submitted 28.11.22; peer-reviewed by R Athe, S Li; comments to author 14.05.23; revised version received 26.05.23; accepted 07.06.23; published 16.08.23.

Please cite as:
Van Dooren M, De Croon R, Swillen A, Verbert K
Bridging the Communication Gap Between People With Cognitive Impairments and Their Caregivers Using mHealth Apps: User-Centered Design and Evaluation Study With People With 22q11 Deletion Syndrome
JMIR Hum Factors 2023;10:e44290
URL: https://humanfactors.jmir.org/2023/1/e44290
doi:10.2196/44290
PMID:37585257

©Martijn Van Dooren, Robin De Croon, Ann Swillen, Katrien Verbert. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 16.08.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Acceptance and Usability of an Innovative mDentistry eHygiene Model Amid the COVID-19 Pandemic Within the US National Dental Practice-Based Research Network: Mixed Methods Study

Jin Xiao¹, DDS, PhD; Dorota Kopycka-Kedzierawski¹, MPH, DDS; Patricia Ragusa¹, BA; Luis Alberto Mendez Chagoya¹, DDS; Kimberly Funkhouser², BSc; Tamara Lischka², BSc; Tong Tong Wu³, PhD; Kevin Fiscella⁴, MPH, MD; Kumari Saswati Kar⁵, DDS; Nisreen Al Jallad¹, DDS, MS; Noha Rashwan¹, DDS; Johana Ren⁵, BSc; Cyril Meyerowitz¹, MS, DDS; National Dental Practice-Based Research Network (PBRN) Collaborative Group⁶

¹Eastman Institute for Oral Health, University of Rochester, Rochester, NY, United States
²Kaiser Permanente Center for Health Research, Portland, OR, United States
³Department of Biostatistics and Computational Biology, University of Rochester, Rochester, NY, United States
⁴Department of Family Medicine, University of Rochester, Rochester, NY, United States
⁵River Campus, University of Rochester, Rochester, NY, United States
⁶School of Dentistry, University of Alabama at Birmingham, Birmingham, AL, United States

Corresponding Author:
Cyril Meyerowitz, MS, DDS
Eastman Institute for Oral Health
University of Rochester
625 Elmwood Ave
Rochester, NY, 14618
United States
Phone: 1 585 275 4925
Email: cyril.meyerowitz@urmc.rochester.edu

Abstract

Background: Amid the COVID-19 pandemic and other possible future infectious disease pandemics, dentistry needs to consider modified dental examination regimens that render quality care and ensure the safety of patients and dental health care personnel (DHCP).

Objective: This study aims to assess the acceptance and usability of an innovative mDentistry eHygiene model amid the COVID-19 pandemic.

Methods: This pilot study used a 2-stage implementation design to assess 2 critical components of an innovative mDentistry eHygiene model: virtual hygiene examination (eHygiene) and patient self-taken intraoral images (SELFIE), within the National Dental Practice-Based Research Network. Mixed methods (quantitative and qualitative) were used to assess the acceptance and usability of the eHygiene model.

Results: A total of 85 patients and 18 DHCP participated in the study. Overall, the eHygiene model was well accepted by patients (System Usability Scale [SUS] score: mean 70.0, SD 23.7) and moderately accepted by dentists (SUS score: mean 51.3, SD 15.9) and hygienists (SUS score: mean 57.1, SD 23.8). Dentists and patients had good communication during the eHygiene examination, as assessed using the Dentist-Patient Communication scale. In the SELFIE session, patients completed tasks with minimum challenges and obtained diagnostic intraoral photos. Patients and DHCP suggested that although eHygiene has the potential to improve oral health care services, it should be used selectively depending on patients’ conditions.

Conclusions: The study results showed promise for the 2 components of the eHygiene model. eHygiene offers a complementary modality for oral health data collection and examination in dental offices, which would be particularly useful during an infectious disease outbreak. In addition, patients being able to capture critical oral health data in their home could facilitate dental treatment triage and oral health self-monitoring and potentially trigger oral health–promoting behaviors.

(JMIR Hum Factors 2023;10:e45418) doi:10.2196/45418
KEYWORDS
teledentistry; mDentistry; oral diseases; virtual visit; intraoral camera; COVID-19; pandemic response; mobile phone

Introduction

Background
Amid the COVID-19 outbreak, dental health care personnel (DHCP) are at a high risk of contracting SARS-CoV-2 because of the close physical proximity between the DHCP and their patients and the absence of enhanced levels of personal protective equipment (PPE) [1-3]. Traditional dental examination relies on person-to-person examination, which poses tremendous challenges during the infectious disease outbreak for reasons including but not limited to infection control, exhaustion of PPEs, chairside time management, and treatment compliance. Although dentistry has been practiced for years using person-to-person visual and tactile intraoral examination now more than ever, dentistry should consider augmenting existing practices with virtual dental services involving a wide variety of technologies and tactics.

Supplementing traditional dental examinations (eg, comprehensive and hygiene recall examinations) with virtual examinations could potentially reduce the exposure risk for patients and DHCPs and preserve a large volume of PPEs that may be in short supply during a pandemic. In the current dental examination model, using the hygiene examination as an example, a single hygiene examination consumes 2 PPEs for the dentist alone because of the need to change PPEs between the dentist’s chairside patient and hygiene examination patient [4]. Traditional hygiene examinations also increase the challenge of infection control because of frequent switching of PPEs and dentists running between dental operatories [4]. In the era of the COVID-19 pandemic, dentistry would benefit from modifying dental examination regimens that render quality care and ensure the safety of patients and DHCP, especially the hygiene examinations.

Objectives
In this digital era, our long-term goal is to develop an innovative mDentistry model (mDent) [5-7]. The mDent leverages the advantages of virtual dental visits and digital mobile health (mHealth) tools, such as intraoral cameras, to deliver virtual oral examinations, treatment planning, and interactive oral health management on a broad population basis [5]. In the mDent model, patients capture intraoral pictures at home before visiting the dentist. Capable patients could perform this independently by watching a photo-taking tutorial video, reducing DHCP instruction time during a virtual visit. The DHCP could assess dental health from intraoral pictures. The dental hygienist would take intraoral x-rays and additional intraoral pictures capturing critical soft and hard tissue in the oral cavity during the hygiene visit. After a convenient virtual dental visit with the dentist to examine findings and treatment plans, patients will have an in-office visit to confirm the findings of the virtual examinations and receive a definite dental treatment, if needed. The conversion of the traditional dental examinations to mDent virtual examinations builds upon the diagnostic reliability of teledentistry [8-10] and the rapid advancement of mHealth tool use by all-age Americans [11,12]. Clinicians have been using intraoral photos [8-10] and live video [13,14] to diagnose caries and predict with high accuracy the appropriate treatment modality for pediatric patients. However, 2 critical components of the mDent model—virtual hygiene examination (eHygiene) and patient self-taking intraoral images (SELFIE)—have not been previously evaluated.

Therefore, in this mDent eHygiene study, we piloted both eHygiene and SELFIE components within the National Dental Practice-Based Research Network (PBRN) [15]. We aimed to assess the feasibility and acceptability of implementing mDent eHygiene while exploring the ability of patients to take internal photos using health tools in their home setting.

Methods

Overall Study Design
This study used a 2-stage implementation investigation to assess the acceptance of 2 components (eHygiene and SELFIE) of the mDent eHygiene model among patients and DHCP (dentists and dental hygienists who were members of the National Dental PBRN). This mDent eHygiene study used mixed methods (quantitative and qualitative) to collect outcome measures and conduct data analysis. Our study protocol has been detailed previously [7]. The study flow is shown in Figure 1.

Figure 1. Specific aims, study design, and outcome measures. SUS: System Usability Score; DPC: Dentist-Patient Communication.

Briefly, the first stage, the eHygiene session, was designed to assess the acceptance of and barriers to mDent eHygiene among patients and DHCP. We enrolled 85 adult patients and 24 DHCP (12 dentists and 12 hygienists) from 12 dental practices in the Northeast region of the National Dental PBRN. The hygienist at each participating practice enrolled approximately 12 (6-15)
hygiene recall patients. A 20-minute instructional video for taking intraoral images was provided to the participating hygienists for training purposes. Patients received 1 in-office hygiene visit to collect the required clinical parameters and intraoral images with the hygienist. These patients then received 1 virtual visit with the dentist to review examination findings and treatment plans.

The second stage, the SELFIE session, was designed to assess the patient’s capability to generate intraoral images using mHealth tools. Hygienists invited one of their patients who completed the first stage eHygiene session to participate in the SELFIE session. In total, 4 patients volunteered to participate in the self-taken intraoral photo session under virtual guidance from the same hygienist during the eHygiene stage.

Outcome Measures and Statistical Analysis

System Usability Scale

The System Usability Scale (SUS) instrument was used to assess the acceptance of the mDent eHygiene approach. The SUS instrument [16-18] is widely adopted in business and technology industries and mHealth fields to measure and quantify the perception of product and service usability. An SUS score >68 indicates an above-average usability [19]. The SUS score of all patients and the dentists and hygienists after each patient visit was calculated. A linear mixed effects model was used to examine factors that influence the SUS score of patients, including patient factors (demographic, socioeconomic, education, order of the patient seen in the eHygiene study, and time spent on eHygiene) while considering the nested random effects within practices and providers. Similar linear mixed effects models were used to examine factors that influence the SUS score of dentists and hygienists, including patient factors, DHCP factors, order of the patient seen in the eHygiene study, and time spent on eHygiene. The order of the patient being seen in the eHygiene study, with a cutoff whether before or after the seventh patient, was built in all the aforementioned models to assess whether the SUS score by DHCP is associated with a learning curve.

Dentist-Patient Communication

The Dentist-Patient Communication (DPC) scale was used to assess how well the patients understood the planned treatment and the quality of the communication between the patients and dentists using eHygiene. We used a modified questionnaire from a validated Doctor-Patient Communication questionnaire [20] that is often used in the medical field. We used a linear mixed effects model to examine the factors that influence the DPC score of patients while considering the nested random effects within practices and providers.

Qualitative Outcomes

We selected 15 individuals (5 patients, 5 dentists, and 5 hygienists) for 30-minute individual interviews virtually. These 15 individuals included those who rated above and below the average SUS score. The questions during the interview addressed the feedback, perceived challenges, and suggestions for improvement for the mDent eHygiene model. The interviews were standardized using interview guides (Multimedia Appendix 1), audio-recorded, transcribed, and analyzed for thematic content.

For the SELFIE session, the key patient tasks assessed were connecting cameras with the tablet, locating the photo-taking module, using a cheek retractor, taking front-view and posterior teeth photos, and ensuring that photographs were stored in the TeleDent platform. The user performance for key tasks was categorized into 3 levels: cosmetic (minor), critical (requiring assistance to proceed), and severe (resulting in significant delays or frustration). A study dentist also assessed the number of photographs and readable photographs using a photo assessment form.

Sample Size Consideration

We calculated the sample size based on the primary outcome, acceptance of mDent by patients that was calculated from patients’ SUS. Various studies [16,21,22] that have used the SUS scale to assess the usability of medical services or mHealth tools report a SUS score with a mean of 47.5-81.2 and an SD of 9.9-21.1. As patients in the eHygiene study were clustered by practice, we used a cluster randomized design for sample size calculation. Assuming that the SUS score difference between the patient-evaluated eHygiene model and other published mHealth tools has a mean of 8 and an SD of 10, a sample size of 72 patients from 12 practices (6 per practice) would achieve 90% power, with an α=0.05. Consideration of sample size for primary and secondary outcomes was detailed in our protocol study [7].

Ethics Approval

This study was peer-reviewed and funded by the National Institute of Dental and Craniofacial Research, United States. This study received single institutional review board approval from the University of Alabama at Birmingham (#300006506) and a local context review from the University of Rochester (#6077).

Results

Overview

The eHygiene study recruited 12 hygienists and 12 dentists from 12 US dental practices located in New York, New Jersey, Pennsylvania, and Connecticut. However, 3 hygienists and 3 dentists withdrew before receiving training because of limited time availability and concerns about the time needed for study activities. Moreover, 1 hygienist and 1 dentist received study training but did not start enrolling patients because of schedule conflicts. A total of 85 patients were recruited and enrolled by participating dental clinics; 2 patients withdrew before receiving intraoral images capture and virtual visits and 1 patient completed the office visit but not the virtual visit.

Usability of eHygiene Among Patients

The demographic characteristics of patients are presented in Table 1. The mean age of the patients was 44.6 (SD 16.2, range 18-74) years. Approximately 70% (58/83) of these patients were female. Most of the patients were White (75/83, 90%), had private insurance (55/83, 66%), and resided in a suburban neighborhood (63/83, 76%). Approximately half of the patients...
had a bachelor’s degree or higher. Interestingly, with all participants owning a smartphone, only 52% (43/83) had used medical care apps; however, none had ever used dental care apps. In addition, 41% (34/83) of the patients had previous experience taking photos of their teeth or mouth with their phones.

Table 1. Demographic characteristics of patients in eHygiene.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Patients (n=83)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD; range)</td>
<td>44.6 (16.2; 18-74)</td>
</tr>
<tr>
<td>Sex (female), n (%)</td>
<td>58 (70)</td>
</tr>
<tr>
<td>Race, n (%)</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>7 (8)</td>
</tr>
<tr>
<td>White</td>
<td>75 (90)</td>
</tr>
<tr>
<td>Other</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Hispanic, n (%)</td>
<td>6 (7)</td>
</tr>
<tr>
<td>Dental insurance, n (%)</td>
<td></td>
</tr>
<tr>
<td>No insurance</td>
<td>17 (20)</td>
</tr>
<tr>
<td>Private</td>
<td>55 (66)</td>
</tr>
<tr>
<td>Government</td>
<td>6 (7)</td>
</tr>
<tr>
<td>Other</td>
<td>6 (7)</td>
</tr>
<tr>
<td>Education, n (%)</td>
<td></td>
</tr>
<tr>
<td>High school</td>
<td>12 (14)</td>
</tr>
<tr>
<td>Some college or associate degree</td>
<td>29 (35)</td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td>25 (30)</td>
</tr>
<tr>
<td>Graduate degree</td>
<td>17 (20)</td>
</tr>
<tr>
<td>Community, n (%)</td>
<td></td>
</tr>
<tr>
<td>Urban</td>
<td>8 (10)</td>
</tr>
<tr>
<td>Suburban</td>
<td>63 (76)</td>
</tr>
<tr>
<td>Rural</td>
<td>12 (14)</td>
</tr>
<tr>
<td>Household income (US $), n (%)</td>
<td></td>
</tr>
<tr>
<td>25,001-50,000</td>
<td>15 (18)</td>
</tr>
<tr>
<td>50,001-100,000</td>
<td>24 (29)</td>
</tr>
<tr>
<td>&gt;100,000</td>
<td>28 (34)</td>
</tr>
<tr>
<td>Prefer not to answer</td>
<td>16 (19)</td>
</tr>
<tr>
<td>Number of household members, n (%)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>11 (13)</td>
</tr>
<tr>
<td>2</td>
<td>28 (34)</td>
</tr>
<tr>
<td>≥3</td>
<td>44 (53)</td>
</tr>
<tr>
<td>Owning a smartphone, n (%)</td>
<td>83 (100)</td>
</tr>
<tr>
<td>Use medical care apps, n (%)</td>
<td>43 (52)</td>
</tr>
<tr>
<td>Use dental care apps, n (%)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Previous experience of taking teeth or mouth photo, n (%)</td>
<td>34 (41)</td>
</tr>
</tbody>
</table>

The eHygiene examination model was well accepted by patients, with a mean SUS score of 70 (SD 23.7). The patients’ SUS score was affected by the time spent on virtual visits and sex (Table S1 in Multimedia Appendix 2). Patients had low SUS scores when they spent more time on the virtual visit \((P=0.003)\). Females reported higher SUS scores \((P=0.03)\). The ratings of the individual items of the SUS are shown in Figure 2A. The response to each SUS item was converted to a scale from strongly negative to strongly positive. Across the 10 items, the responses from more than 60% (50/83) of the participants were neutral or positive (positive and strongly positive), indicating the well-perceived usability of the eHygiene model. For
example, 60% (50/83) of the participants thought the eHygiene examination model was easy to use; 12% (14/83) of the participants felt that they needed technical support to use the eHygiene examination.

**Figure 2.** Acceptance of eHygiene among patients, hygienists, and dentists assessed by the System Usability Scale.

---

**Usability of eHygiene Among DHCP**

Overall, the race of 55% (5/9) of the dentists and 66% (6/9) of the hygienists was White. Half of the dentists and all hygienists were female. The average age was 56 (SD 7.0) years for the dentists and 42 (SD 12.5) years for the hygienists. The eHygiene model was moderately accepted by dentists (SUS score: mean 51.3, SD 15.9) and hygienists (SUS score: mean 57.1, SD 23.8). After examining each patient, every dentist or hygienist provided an SUS score for the eHygiene model, resulting in multiple SUS score values for each practitioner. The collected SUS scores from dentists and hygienists exhibited an association with a learning curve, wherein lower SUS scores were observed among...
patients treated initially, especially among the first 7 patients included in the study (dentist: \( P = .07 \); hygienist: \( P = .06 \); Figure 3). The ratings of the individual items of the SUS are shown in Figure 2B for the hygienists and Figure 2C for the dentists. Among the 2 items that are related to the learnability structure of the SUS [23], item 4 (I think that I would need the support of a technical person to be able to use eHygiene) and item 10 (I needed to learn a lot of things before I could get going with eHygiene), the results further suggested a learning curve related SUS learnability score among hygienists (Figure S1 in Multimedia Appendix 2) and dentists (Figure S2 in Multimedia Appendix 2). Furthermore, dentists reported lower SUS scores when spending more time on the virtual visit (\( P = .04 \)). The results of the linear mixed effects regression model are detailed in Table S1 in Multimedia Appendix 2.

Figure 3. Learning curve of dentists and hygienists. Dentists’ (A) and hygienists’ (B) System Usability Scale (SUS) scores appear to be associated with a learning curve, with a higher SUS score given to patients who were seen after the first 7 patients in the study, after adjusting for dentist and hygienist demographics, patient characteristics, and time spent conducting hygiene examination visits (dentist: \( P = .07 \); hygienist: \( P = .06 \)).
DPC Using eHygiene

No statistical difference was observed between the current hygiene examination DPC (mean 58.5, SD 3.8) and the eHygiene examination DPC (mean 58.1, SD 5.97; \( P = .51 \)). Both dentists and patients expressed that the oral findings and planned treatment were well understood by the patients (Table 2). The linear mixed effects regression model revealed that the current hygiene examination DPC was higher \(( P = .01)\) among patients residing in suburban communities after adjusting for other patient factors listed in Table S2 in Multimedia Appendix 2. Interestingly, the eHygiene examination of DPC was influenced by patient ethnicity. Hispanic patients had lower DPC scores than non-Hispanic patients \(( P = .01)\).

### Table 2. Dentist and patient communication in the current hygiene model and the eHygiene model.

<table>
<thead>
<tr>
<th>Dentist-Patient Communication score (patient evaluated, maximum score=60), mean (SD)</th>
<th>Current hygiene examination</th>
<th>eHygiene examination</th>
<th>( P ) value (( t ) test)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dentist-Patient Communication score</strong> (patient evaluated, maximum score=60), mean (SD)</td>
<td>58.5 (3.8)</td>
<td>58.1 (5.97)</td>
<td>.51</td>
</tr>
<tr>
<td><strong>How well do you think your patient understood what you explained about their oral health and the treatment you recommended?</strong> (Dentist evaluated, eHygiene n=83 patient visits), n (%)</td>
<td></td>
<td></td>
<td>N/A(^a)</td>
</tr>
<tr>
<td>Very well</td>
<td>4 (44)</td>
<td>46 (55)</td>
<td></td>
</tr>
<tr>
<td>Fairly well</td>
<td>5 (56)</td>
<td>36 (43)</td>
<td></td>
</tr>
<tr>
<td>Poor</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Very poor</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)N/A: not applicable.

Patients Capable of Taking Diagnostic Intraoral Images

In the SELFIE session, all patients were able to complete the tasks with no or minor challenges (Table S3 in Multimedia Appendix 2). The longest time was spent on taking posterior teeth photographs (mean 4.7, SD 1.5 min). All other tasks, including capturing images of the front teeth and uploading photographs to TeleDent, took approximately 1 minute. No difference was found between patients and hygienists in terms of the total number and diagnostic qualities of the images taken (Table S4 in Multimedia Appendix 2). For instance, patients took an average of 26.5 photographs per person, whereas hygienists took an average of 33.3 photographs per person \(( P = .66)\). More than half of the photographs taken by the patients were diagnostic, indicating that the photographs were clear and included all anatomical structures of the teeth for diagnosis.

Perception of the eHygiene Model by Patients and DHCP

Tables 3 and 4 indicate the quantitative perceptions of the current hygiene model and eHygiene model by patients and DHCP. The average time for taking intraoral images was 10 minutes. Most of the patients (76/83, 91%) reported that the photograph-taking procedure was comfortable. Dentists and patients consistently reported that the average time spent on a virtual visit was 6 minutes. Overall, the participating hygienists and the dentists thought the eHygiene model was suitable for the majority of patients with good oral health who did not have restorative or periodontal treatment in the past 1 year or more. Most of the DHCP indicated that eHygiene might not be suitable for patients with poor oral health who had ongoing untreated caries or periodontal disease, or for patients with urgent oral needs; for example, pain and orofacial swelling.
Table 3. Perspective on current (in-person) hygiene examination model by patients, hygienists, and dentists.

<table>
<thead>
<tr>
<th>Survey questions</th>
<th>Patients’ perspective (n=83)</th>
<th>Hygienists’ perspective (n=9)</th>
<th>Dentists’ perspective (n=9)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Time for in-person hygiene examination (min), n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-4</td>
<td>23 (28)</td>
<td>3 (33)</td>
<td>1 (11)</td>
</tr>
<tr>
<td>5-10</td>
<td>55 (66)</td>
<td>6 (67)</td>
<td>8 (89)</td>
</tr>
<tr>
<td>&gt;10</td>
<td>5 (6)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Time waiting for the dentist for examination (min), n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-4</td>
<td>54 (65)</td>
<td>6 (67)</td>
<td>N/A</td>
</tr>
<tr>
<td>5-10</td>
<td>26 (31)</td>
<td>3 (33)</td>
<td>N/A</td>
</tr>
<tr>
<td>&gt;10</td>
<td>3 (4)</td>
<td>0 (0)</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>PPE(^b) changed when switching from seeing the chair side patient to a hygiene recall examination patient, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gloves</td>
<td>N/A</td>
<td>N/A</td>
<td>9 (100)</td>
</tr>
<tr>
<td>Surgical mask</td>
<td>N/A</td>
<td>N/A</td>
<td>5 (56)</td>
</tr>
<tr>
<td>N95 mask</td>
<td>N/A</td>
<td>N/A</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Gown</td>
<td>N/A</td>
<td>N/A</td>
<td>1 (11)</td>
</tr>
<tr>
<td>Face shield or goggles</td>
<td>N/A</td>
<td>N/A</td>
<td>2 (22)</td>
</tr>
<tr>
<td>Bonnet</td>
<td>N/A</td>
<td>N/A</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Charged PPE fees for hygiene recall examination patients since the COVID-19 pandemic started, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>N/A</td>
<td>N/A</td>
<td>7 (78)</td>
</tr>
<tr>
<td>1%-25%</td>
<td>N/A</td>
<td>N/A</td>
<td>0 (0)</td>
</tr>
<tr>
<td>26%-50%</td>
<td>N/A</td>
<td>N/A</td>
<td>1 (11)</td>
</tr>
<tr>
<td>51%-75%</td>
<td>N/A</td>
<td>N/A</td>
<td>0 (0)</td>
</tr>
<tr>
<td>76%-100%</td>
<td>N/A</td>
<td>N/A</td>
<td>1 (11)</td>
</tr>
<tr>
<td><strong>Do you think it would be helpful if dental professionals used images on the computer or tablet to explain your oral health?, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes, very helpful</td>
<td>40 (48)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Yes, to some degree</td>
<td>40 (48)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>No, not helpful</td>
<td>0 (0)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>I do not know</td>
<td>3 (4)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Do you routinely take intraoral images for your patients during hygiene visits?, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Always</td>
<td>N/A</td>
<td>2 (22)</td>
<td>1 (11)</td>
</tr>
<tr>
<td>Very often</td>
<td>N/A</td>
<td>0 (0)</td>
<td>3 (33)</td>
</tr>
<tr>
<td>Sometimes</td>
<td>N/A</td>
<td>1 (11)</td>
<td>2 (22)</td>
</tr>
<tr>
<td>Rarely</td>
<td>N/A</td>
<td>1 (11)</td>
<td>3 (33)</td>
</tr>
<tr>
<td>Never</td>
<td>N/A</td>
<td>5 (56)</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Do you use oral or teeth images to facilitate patient education or treatment planning?, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Always</td>
<td>N/A</td>
<td>2 (22)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Very often</td>
<td>N/A</td>
<td>2 (22)</td>
<td>3 (33)</td>
</tr>
<tr>
<td>Sometimes</td>
<td>N/A</td>
<td>4 (44)</td>
<td>4 (44)</td>
</tr>
<tr>
<td>Rarely</td>
<td>N/A</td>
<td>0 (0)</td>
<td>2 (22)</td>
</tr>
<tr>
<td>Never</td>
<td>N/A</td>
<td>1 (11)</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

\(^a\)N/A: not applicable (the survey question was not answered by the group of participants).

\(^b\)PPE: personal protective equipment.
Table 4. Perspective on the eHygiene examination model by patients, hygienists, and dentists.

| Survey questions and category | Patients’ perspective (n=83) | Hygienists’ perspective (n=83 patient visits, assessed by 9 hygienists) | Dentists’ perspective (n=83 patient visits, assessed by 9 dentists) |
|———|———|———|———|
| Perceived patient’s comfortability when intraoral images were taken, n (%) | | | |
| Very comfortable | 60 (72) | 19 (23) | N/A<sup>b</sup> |
| Somewhat comfortable | 16 (19) | 38 (46) | N/A |
| Somewhat uncomfortable | 5 (6) | 18 (22) | N/A |
| Very uncomfortable | 0 (0) | 5 (6) | N/A |
| Extremely uncomfortable | —<sup>c</sup> | 1 (1) | N/A |
| Which types of patients do you think should be considered for the eHygiene virtual visits? (Choose all that apply), n (%) | | | |
| None | N/A | 0 (0) | 12 (14) |
| Patients with good oral health who did not have restorative or periodontal treatment in the past 1+ year | N/A | 62 (75) | 70 (84) |
| Patients with poor oral health who had ongoing untreated caries or periodontal disease. | N/A | 9 (11) | 0 (0) |
| Patients with nonurgent oral diseases (eg, caries, periodontal pocket deeper than 4 mm, etc) identified by hygienists during cleaning. | N/A | 32 (39) | 2 (2) |
| Patients with urgent oral needs, for example, pain, orofacial swelling, etc. | N/A | 7 (8) | 1 (1) |
| Patients with oral mucosal lesions identified by hygienists during cleaning. | N/A | 23 (28) | 0 (0) |
| All patients | N/A | 7 (8) | 0 (0) |

<sup>a</sup>The eHygiene model was well accepted by patients (System Usability Scale score: mean 70.0, SD 23.7) and moderately accepted by hygienists (System Usability Scale score: mean 57.1, SD 23.8) and dentists (System Usability Scale score: mean 51.3, SD 15.9). Time spent (min) on taking intraoral images by hygienists (mean 9.8, SD 5.7; range 1-25). Time spent (min) conducting the eHygiene virtual visit by patients (mean 5.9, SD 5.3; range 1-30) and dentists (mean 6.4, SD 5.1; range 2-40). Of the 83 patient visits, eHygiene visits during work hours by patients was 41 (49%) and for dentists 29 (35%).

<sup>b</sup>N/A: not applicable (the survey question was not applicable to the specific group of participants).

<sup>c</sup>Not available.

The qualitative analysis further indicates 8 thematic patterns of patients’ and DHCP’s perspectives on the eHygiene model. The first 4 are the foreseen benefits of eHygiene, which include an overall positive experience for patients (theme 1), eHygiene enables effective communication to the patient about oral health (theme 2), eHygiene saving resources (theme 3), and eHygiene has the potential to improve oral health care services (theme 4). However, patients and DHCP also talked about 4 limitations of eHygiene: the current eHygiene model does not provide all necessary oral health data needed to make comprehensive evaluations (theme 5), extra time is needed because of the technology-related learning curve and technical issues (theme 6), eHygiene lacks interpersonal interaction (theme 7), and selectivity in eHygiene use (theme 8). Representative quotes are listed in Table S5 in Multimedia Appendix 2.

In addition, the participating dentists and hygienists raised concerns regarding monetary and reimbursement issues. Specifically, they pointed out that the time required for intraoral photo capture and virtual dental visits was not traditionally reimbursed under the current dental fee schedule. As such, they emphasized the need for recognition of these services by insurance payers or patients to facilitate adoption of the service. Addressing these concerns and finding viable reimbursement options may help to increase the adoption and sustainability of the mDent model.

Discussion

Principal Findings

The eHygiene study had multiple strengths. First, we were successful in engaging patients and nondentist professionals in the dental office in capturing essential oral health data (intraoral images and x-rays) and conducting virtual dental hygiene examinations (mDent model). This inclusive engagement is novel and potentially transformative to dental practice. The use of smartphones and mobile devices to take photos of the mouth and teeth and conduct oral disease screening has been recently reported [6,24,25]; however, the feasibility of engaging dental hygienists and dental patients to obtain intraoral images using an intraoral camera has not been previously assessed. Second, transforming the traditional one-to-one visual and tactile dentist examination to an eHygiene visit requires a team effort from several stakeholders, including patients, dental hygienists, and dentists. The team efforts could lead to better DHCP-patient communication and a better understanding of and compliance with this approach to dental treatment. Third, integrating the...
mHealth concept into dentistry to achieve population-wide oral health screening and monitoring is extremely innovative and offers a vehicle to promote patient-engaged oral health education and early detection of patient-driven oral diseases. Fourth, the eHygiene model is a novel way of preserving PPEs during the COVID-19 pandemic and other respiratory transmissible disease outbreaks. In addition to the abovementioned strengths, the eHygiene study has the following limitations: (1) a limited number of patients participated in the SELFIE session, which limits the generalization of the study results from the SELFIE session; and (2) the study was conducted in dental clinics residing in suburban areas with most of the participating patients having private dental insurance. Therefore, the study results cannot be generalized to the underserved population or dental office in the community setting.

During the early time of the COVID-19 outbreak, the American Dental Association issued recommendations for their fellow dentists to provide care to emergency patients only. According to the American Dental Association, as of 2019, general dentists in the United States are delivering 564 million patient visits per year [26]. Notably, 316 million (56%) of these 564 million patient visits are examination visits that are often not linked to definitive treatment delivery at the same visit. Under the circumstances that routine office dental visits were reduced during the COVID-19 outbreak, there were anecdotal reports that some patients and dentists chose to use intraoral images taken by patients to assess urgent dental situations; for example, fractured tooth and facial swelling. This new phenomenon of patient-dentist communication provides an opportunity for a new way of delivering dental service, mDentistry, that could transform community dentistry.

Although the eHygiene model was well accepted by patients, the eHygiene model was conservatively accepted by dentists and dental hygienists. On the basis of the study findings, it appears that the SUS scores of both dentists and hygienists may correlate with a learning curve. Specifically, higher SUS scores were observed in visits that occurred after the first 7 patients were seen in the study. This suggests that with experience and practice, dental professionals may become more comfortable and proficient in using the system, leading to increased user satisfaction. Furthermore, SUS measures both usability (8 items) and learnability (2 items: item 4 and item 10) [23]. The subanalysis of the SUS learnability of the dentists and hygienists confirmed a potential learning curve associated with the eHygiene model. This indicates that additional training may be necessary before DHCP implement the eHygiene model in their daily clinical practice. Another intriguing point raised by dentists and hygienists was monetary and reimbursement concern, which might also be associated with the SUS score given by DHCP. Although some hygienists expressed that patients would save time and money by using eHygiene, others expressed concerns about losing revenue or not being appropriately reimbursed for eHygiene services because of the extended appointment time. Insurance reimbursement for eHygiene work was considered crucial to support the sustainability of the mDentistry eHygiene model.

In addition, the eHygiene study used SUS to assess acceptance, and the other system that has been used to assess individuals’ acceptance of technology service is the UTAUT (Unified Theory of Acceptance and Use of Technology) [27-29]. Although SUS is a widely used questionnaire-based tool for measuring the perceived usability of a system or product; the UTAUT, on the other hand, is a theoretical model that seeks to explain and predict user acceptance and use of technology [30-32]. The UTAUT is useful for predicting and understanding user acceptance and use of technology in various contexts and can be used to identify potential barriers to technology adoption and use [33]. Both frameworks can be used to improve the design and development of technology-based products and systems.

The mDentistry eHygiene model offers a complementary modality for oral health data collection, which would be particularly useful during an infectious disease outbreak. The fact that patients can capture critical oral health data in a home setting could facilitate dental treatment triage and oral health self-monitoring and potentially trigger oral health–promoting behaviors. This hypothesis of self-monitoring associated with oral health behavior changes could be tested in a clinical trial that assessed the impact of mDent on oral health promotion at a population base.

Notably, mDentistry is at the intersection of incorporating artificial intelligence (AI) into dentistry. AI represents an emerging adjunct to caries screening and risk management, building upon (1) reliability of teledentistry that uses intraoral images and live videos to make diagnostic decisions [9,13,14,34] and (2) rapid advancement of mHealth tools use by all-age Americans [35,36]. Recently, AI had been tested in detecting caries and oral pathologies on dental x-rays [37,38]. Our team has developed a smartphone app, AICaries, that uses AI-powered technology to detect caries on photographs of teeth taken via smartphone [39-41]. As AI is currently used to aid imaging recognition to improve disease diagnosis in many medical fields, including oncology, ophthalmology, and radiology [42-45], AI has the full potential to be developed in dentistry for remote caries detection and caries risk management for underserved patients with limited access to oral health care. Future clinical service transformation should leverage the convenience provided by mDentistry and the robust disease screening powered by AI technology to improve oral health early detection and prevention at a broad population base. A population-wide intraoral images and x-rays database is urgently needed to be developed to facilitate oral disease screening automation in the community.

Conclusions

The eHygiene study results informed the process and usability of the mDentistry eHygiene model and showed promise for conducting virtual dental examinations and empowering patients with mHealth tools. The eHygiene model was well accepted by patients and moderately accepted by dentists and hygienists. Changes in reimbursement could accelerate its adoption. In addition, the complementary modality for oral health data collection and examination in dental offices provided by eHygiene would be beneficial during an infectious disease outbreak.
Acknowledgments
This study was funded by National Institutes of Health grants X01DE030396, U19-DE-28717, and U01-DE-28727. The opinions and assertions contained herein are those of the authors and should not be construed as necessarily representing the views of the respective organizations or the National Institutes of Health.

Authors’ Contributions
JX, CM, DK-K, TTW, and KF contributed to the conception, design, data acquisition, analysis, interpretation, and drafting and critical revision of the manuscript. PR, KF, TL, KSK, NAJ, NR, JR, and the National Dental Practice-Based Research Network (PBRN) Collaborative Group contributed to data acquisition, analysis, and critical revision of the manuscript. All authors have read and approved the final version of the manuscript and agree to be accountable for all aspects of this work. Details about the National Dental PBRN can be found on the internet [46].

Conflicts of Interest
None declared.

Multimedia Appendix 1
Interview guide.
[DOCX File, 37 KB - humanfactors_v10i1e45418_app1.docx ]

Multimedia Appendix 2
Supplementary material.
[DOCX File, 243 KB - humanfactors_v10i1e45418_app2.docx ]

References
lower-income-americans-are-becoming-increasingly-dependent-on-smartphones-says-study/?sh=70db17961083 [accessed 2022-11-08]


https://humanfactors.jmir.org/2023/1/e45418
Abbreviations

AI: artificial intelligence
DHC: dental health care personnel
DPC: Dentist-Patient Communication
mHealth: mobile health
PBRN: Practice-Based Research Network
PPE: personal protective equipment
SUS: System Usability Scale
UTAUT: Unified Theory of Acceptance and Use of Technology
Original Paper

Primary Perspectives in Meme Utilization as a Digital Driver for Medical Community Engagement and Education Mobilization: Pre-Post Study

Darrel Wang¹, BS; Neha Balapal², BS; Amala Ankem³, MPH; Saishravan Shyamsundar⁴, BS; Adarsh Balaji⁵, BS; Jasmine Kannikal⁶, MIB, MBS; Marlie Bruno⁷; Shuhan He⁸,⁹, MD; Paul Chong¹⁰, BS

¹Arizona State University, Phoenix, AZ, United States
²CUNY School of Medicine, New York, NY, United States
³Lab of Computer Science, Massachusetts General Hospital, Boston, MA, United States
⁴Geisinger Commonwealth School of Medicine, Scranton, PA, United States
⁵American University of the Caribbean School of Medicine, Cupecoy, Netherlands
⁶University of Miami Miller School of Medicine, Miami, FL, United States
⁷Northeastern University, Boston, MA, United States
⁸Department of Emergency Medicine, Massachusetts General Hospital, Boston, MA, United States
⁹Center for Innovation in Digital HealthCare, Massachusetts General Hospital, Boston, MA, United States
¹⁰Campbell University School of Osteopathic Medicine, Lillington, NC, United States

Corresponding Author:
Darrel Wang, BS
Arizona State University
411 N Central Ave
Phoenix, AZ, 85004
United States
Phone: 1 6266361615
Email: dwang150@asu.edu

Abstract

Background: Memes have gone “viral,” gaining increasing prominence as an effective communications strategy based on their unique ability to engage, educate, and mobilize target audiences in a call to action through a cost-efficient and culturally relevant approach. Within the medical community in particular, visual media has evolved as a means to influence clinical knowledge transfer. To this end, the GetWaivered (GW) project has leveraged memes as part of a behavioral economics toolkit to address one of the most critical public health emergencies of our time—the 20-year opioid epidemic. As part of a multidimensional digital awareness campaign to increase Drug Enforcement Administration (DEA)-X waiver course registration, GW investigated the results of meme usage in terms of impressions, website traffic, and ultimately user acquisition, as determined by web-based training enrollment and attendance outcomes.

Objective: The objective of this study was to determine the efficacy of implementing humor-based promotional content versus the traditional educational model, and how the translation of the increase in engagement would increase the participant count and website traffic for GW’s remote DEA-X waiver training.

Methods: The approach to this study was based on 2 time frames (pre- and postcampaign). During April-July 2021, we developed a campaign via advertisements on Facebook, Twitter, Instagram, and the GW website to expand outreach. These memes targeted medical professionals with the ability to prescribe buprenorphine. The time frame of this campaign measured engagement metrics and compared values to preceding months (January-March 2021) for our GetWaivered website and social media pages, which translated to registrants for our remote DEA-X waiver training.

Results: By the end of July 2021, a total of 9598 individuals had visited the GW website. There was an average of 79.3 visitors per day, with the lowest number of daily visitors being 0 and the highest being 575.

Conclusions: The use of memes may provide a medium for social media engagement (likes, comments, and shares) while influencing viewers to pursue a proposed action, such as e-training registration.
buprenorphine; DEA X-waiver; opioid use disorder; OUD; MAT; opioid; waiver; medical education; continuing education; e-training; e-learning; awareness; social media; engagement; marketing; promotion; meme; campaign; advertisement; advertising; outreach; Facebook; Twitter; Instagram

Introduction

Background

In combating the opioid crisis, buprenorphine is the drug of choice when adhering to an evidence-based clinical standard of care. However, the Drug Addiction Treatment Act of 2000 mandates those clinicians receive a waiver known as the Drug Enforcement Administration (DEA)-X waiver in order to prescribe buprenorphine [1]. Due to the many historical accessibility barriers associated with obtaining a DEA-X waiver, only about 5% of physicians in the United States have acquired this prescriptive authority, which, in turn, limits the ability to care for patients with opioid use disorder (OUD) [2]. GetWaivered (GW) was founded to address this caveat by encouraging and facilitating more clinicians in obtaining their DEA-X waiver. Based upon exploratory research, GW’s model focused on behavioral nudges to address identified barriers as to why clinicians do not obtain their “X” waiver. They include (1) the absence of a social norm, (2) hassle bias in obtaining the waiver, and (3) a lack of salience in treating OUD [3]. Additionally, learning barriers were amplified in the context of the COVID-19 pandemic, as social distancing practices further limited access to traditional medical education, including the DEA-X waiver training courses. Therefore, there has been a push toward web-based platforms for medical education. Outreach efforts since the COVID-19 pandemic have also moved toward social media platforms due to the increase in clinician use [4]. A previous digital campaign performed in April-May 2020 showed that 13% of traffic to the GW website came from social media sources (Facebook, Instagram, and Twitter). Evaluating social media engagement can be key to increasing awareness of our course [5].

Twitter, a social media platform, has around 340 million active users, with about 500 million Tweets (microblogs) posted per day in a wide variety of locations and languages [6]. Twitter is one of the most popular forms of social media used for health care communication, which is why more people are continuously using it as a data source for research [7]. Twitter’s ease of rapidly distributing published information, accelerating peer review, and engaging large numbers of people from various communities make it a key player in professional education in health care [7]. Despite the constant evolution of social media platforms, advertising, and marketing, humor has been a staple that has withstood the tests of social media evolution. The engagement that is provoked when humor is at the forefront of any message allows for reader engagement through collective laughter and relatability [8]. A meme can be best described as a photo with a witty caption. Memes are often created to be pop-culturally themed pictures with words that are either captioned or overlaid on the photo. GW was able to target memes created with an emphasis on modern-day findings of the behavioral aspects surrounding the viewership of the artwork via focal viewpoints [9]. The campaign focused on the effectiveness of a text overlay on top of a photo that is widely known in popular culture. The focal viewpoint, in the case of meme content creation, shifts the focus from examining a picture to simultaneously reading an in-figure caption. The pop-culture references, which entailed largely popular pictures, were overlaid with words that provided comical references and statements to encourage reaction formation from audience members. Measures included engagement from the health care community via social media.

Social media is both collaborative and user-generative focused, which promotes participatory learning and action among its users [10]. Existing public engagement with scientific social media is one-directional, but increasing participation through creative posts and discussions can provide new avenues of communicating with target audiences [11]. However, quantifying the impact of social media on health campaigns has proven to be a challenge [12-14]. Overall, web analytics can contribute to determining a website’s usability and conversion rates [15]. Specifically, for evaluating the efficiency of Facebook campaigns for large-scale public health recruitment, such as for COVID-19, cost per click and cost per response are effective outcome measures [16]. Furthermore, while the measurement approaches for measuring social media can be quantitative, qualitative, or mixed methods, it is recommended that the value of social media in a health care endeavor be evaluated by analyzing pre- and post-social media adoption [17].

Data surrounding the efficacy of memes as the content of an outreach campaign have been lacking, so we sought to analyze data on the use of memes to engage social media users in a mission to drive DEA-X waiver registration and, in turn, increase the number of nationwide clinicians that have fulfilled mandated requirements to prescribe buprenorphine. Overall, this study is important for both theory and practice [17], as it contributes to the accumulation of knowledge in the emerging field of health care social media analytics while also enabling the GW team to design social media strategies that yield enrollment results.

Objectives

Our first aim was to create a meme-based social media approach to reach greater audiences while using a digital framework methodology for increases in viewer count and the translated registration numbers for our e-training [18]. This first step resulted in the creation of memes by our content creation team. Additionally, we were able to assess the study’s time frame for participants based on their previous registration numbers. We examined if there were objectively better reaches to our audience and assessed the sustainability and efficacy of the process.
Our second aim was to analyze if the methodology could create more engagement from the community in the form of post likes, shares, and training registrations. The laugh model was used (Table 1) as a method of community outreach that resulted in relatability to pop-culture references to drive social media engagements [19]. The objective of the laugh model was to implement a low-cost method of promotion for public health crises. The framework primarily focused on the communication of public health issues through humor and pop-culture references. The model itself sought to increase awareness, as highlighted on [20]. The conclusions from this study indicate an increase in effectiveness when humor-based promotional measures are compared against education-focused efforts. The laugh model’s part in GW’s desire to increase participation in DEA-X waiver training would allow for the empowerment of humor-driven marketing tactics to increase viewership. The method, in turn, would increase participation in GW’s web-based training for understanding the DEA-X waiver logistics and buprenorphine mechanisms.

Table 1. Use of the laugh model to promote GetWaivered awareness.

<table>
<thead>
<tr>
<th>Framework component</th>
<th>GetWaivered campaign</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consumer motives</td>
<td>• Need for more DEA4-X waived clinicians to help treat opioid use disorder to help curb the opioid epidemic</td>
</tr>
<tr>
<td></td>
<td>• Need to spread awareness on an easy-to-share web-based platform with engaging content</td>
</tr>
<tr>
<td>Priority health needs</td>
<td>Eligible clinicians need to obtain a DEA-X waiver in order to prescribe buprenorphine, an evidence-based treatment for opioid addiction. These clinicians need to be nudged and made aware of the need to become waived</td>
</tr>
<tr>
<td>Message development</td>
<td>Use of memes and other shareable content to encourage people to visit the GetWaivered website</td>
</tr>
<tr>
<td>Web 2.0</td>
<td>Use of Facebook, Twitter, and Instagram platforms; Twitter outperforms the other platforms in reach and engagement and has a larger following.</td>
</tr>
<tr>
<td>Social momentum</td>
<td>Increased content engagement results in a sharing, retweeting, and organic user-generated advertising “domino effect,” with viral content inducing substantial surges in social media traffic.</td>
</tr>
<tr>
<td>Public health impact</td>
<td>Total website visitors (April 1, 2021-July 31, 2021): 9598</td>
</tr>
<tr>
<td></td>
<td>E-training registrants (May 2021-July 2021): 396</td>
</tr>
<tr>
<td>Sustainability</td>
<td>GetWaivered meme creation and implementation expense: US $600 per month, with ongoing social media account maintenance performed voluntarily in-house at no cost.</td>
</tr>
</tbody>
</table>

aDEA: Drug Enforcement Administration.

Methods

Ethics Approval

The study was approved by the Mass General Brigham review board (#2021P000447). The procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2000 (5).

Target Audience

The intended target of the social media postings was clinicians who can prescribe buprenorphine. This included attending physicians, resident physicians, nurse practitioners (NPs), physician assistants (PAs), and health professional school students [21]. These students (NP, PA, and medical) and residents, upon completion of their degree plans, will receive their DEA-X waiver upon the completion of their training, as well as their official license numbers. The evaluated Twitter, Instagram, and Facebook posts were posted between April 2021 and July 2021 and analyzed by metrics including engagement, users, acquisition, content, and platform [15].

Social Media Program, Content, and Promotion

Research supports the incorporation of social media as a viable channel capable of integrating cohesive messaging and furthering mission-driven organizational objectives, particularly as they relate to public health and welfare [12]. GW sought to leverage these insights to empower clinicians in treating OUD through evidence-based interventions via a multidimensional model based on capacity-building.

Project Design

GW first sought to expand outreach efforts through its social media interface before using memes to advocate for DEA-X waiver course enrollment. This biphasic model consisted of 3 primary digital activities: (1) conceptualization of organic content with a unified call-to-action, (2) deployment of this material through a series of timed, course-specific social media advertisements to motivate follow-through, and (3) real-time monitoring of course registration through the course registration page [22] to inform strategic responsiveness in terms of creative direction and distribution methodology. The first stage of the model focused on the generation of memes, based on pop-culture references, and the dissemination of information via social media. The second stage of the model focused on the comparison of numerical values of social media and website engagers, to previously measured numbers prior to humor-based promotions. The target audience was clinicians with current or prospective eligibility to prescribe buprenorphine. This included attending and resident physicians, NPs, and PAs, as well as students in each of these respective cohorts [21]. Residents and students were incentivized to complete GW’s waiver course preemptively to streamline the waiver acquisition process upon degree completion and conferral of licensure. The variables of interest, including but not limited to engagement, user acquisition, content-associated metrics, and overall platform data, were...
measured during the 4-month period from April 2021 to July 2021 from GW’s Twitter, Instagram, and Facebook accounts [15].

**Social Media Programming: Content Conceptualization, Delivery, and Promotion**

Based on the laugh model (Table 1), humor-embedded informational content was presented to an audience segmented according to prevalent attitudes toward buprenorphine use in the treatment of OUD. This included those who (1) were unaware of the intervention, (2) were in favor but unmotivated to get DEA-X waivered, (3) were undecided about its adoption, and (4) expressed opposition. For those in category 1, the main goal of message development was to educate; this consisted of presenting objective information about buprenorphine, its administration, and its proven clinical efficacy. Regarding those in category 2, an emphasis was placed on the support GW provides to minimize the hassle and bias associated with obtaining the waiver. Appealing to those in category 3 involved persuasive tactics directed toward altruistic drivers of behavior and a sense of solidarity with peers and role models within the medical community. Lastly, engaging those in category 4 depended on rebuttals that challenged common arguments through myth-busting graphics and captions. To maximize target audience internalization, time-relevant memes (Figure 1) were leveraged as a vehicle for this messaging, to exploit viral trends and influence perception as less confrontational or argumentative.

GW’s meme content, in addition to flyers, was delivered in regularly spaced intervals to avoid overwhelming followers with a flux of posts all at once, or a lack thereof, resulting in user attrition. The goal of this social media schedule was 2-fold: to increase the number of followers through a consistent digital presence and promote GW website traffic to capitalize on another opportunity to influence primed behaviors (eg, waiver course enrollment). Although the decision-making stage of the social media user affects the likelihood of an endorsed action, consecutive reinforcement via multiple digital touch points can exponentially increase this probability (Figure 2).

Content promotion on Facebook, Instagram, and Twitter was executed through both direct and indirect strategies. The direct promotion was achieved primarily through advertisement investments on Facebook and Instagram. The content was “boosted” among a custom-selected audience that varied by geographical location, clinical position, interests, and demographic variables. Furthermore, the promotional campaign duration was defined according to the days between postlaunch and the next e-course date to maximize enrollment.

Moreover, it is important to note that these efforts were not mutually exclusive; due to the integration of social media, GW-branded content was cross-promoted across platforms simultaneously. For instance, in the process of making a post “live” on Instagram, the option to share it on Facebook was also selected; Facebook users are alerted to the Instagram account source when this content appears on their respective feeds, which has the potential to convert the Facebook user into an Instagram follower as well, as a consequence of a single content consumption experience.

Indirect methodologies depended on a network of recruited influencers within the health care space, further segmented by professional status (eg, a medical student), professional specialty (eg, an NP), or organization (eg, the Florida Medical Association). In addition to tagging high-reach influencers directly on an image, they were also sent direct messages with a gentle request to repost either the image itself, an e-flyer, or an active story presenting the same image or training-related information. Hashtags were also leveraged as a means to increase exposure to GW’s target audience.
Figure 1. Examples of memes.

Figure 2. Meme dissemination framework.

**Mobilization**

The success of our call to action through social means was evaluated based on both registration and participation in our DEA-X waiver training e-course. Of those in attendance, pharmacology and administration were taught through instructional and clinical case-based learning pedagogies by certified experts in addiction medicine. Course instructors then prepared the graduates for the next phase of mobilization by providing a roadmap on how to navigate the administrative aspects of obtaining a DEA-X waiver. The first step, submitting a Notice of Intent form, was done in class, enabling any issues to be addressed instantaneously. Prompting clinicians to initiate the process in class made them more likely to follow through to completion. The training concluded with both local and national resources to support clinicians with any concerns or questions, in addition to a post-course email and our social media linking back to this same information so that it was readily accessible through multiple interfaces.
The flowchart illustrates the self-servicing, repetitive cycle that the promotional meme content is based upon. The success of the cyclical nature of the process is dependent on consumer engagement, which results in course registration for clinicians.

The compilation of memes provides several examples of successful engagement outreach from the GW social media avenues (Twitter, Instagram, and Facebook). With trending pop-culture photos, GW was able to tie visual representations of emotions and actions to the relatable text regarding the efficacy of GW’s web-based DEA-X waiver training.

The pie charts illustrate the demographic breakdown by self-identified gender and the differences between the populations of returning web visitors and first-time web visitors.

Data Collection and Analysis
The assessment of website metrics, including page traffic and course registration, was gathered, and each impression was noted. These measures allowed for a general understanding of the impact that each promotional post had on training registration. In addition, the timelines associated with meme launches coincide with an improvement in outreach, as seen by the analytical trends.

Users are defined as visitors who have started at least one session with a website during a specified period. A session is defined as a group of user interactions with a website. Page views, also known as page impressions, are defined as the total number of views a website has had. Google Analytics defines direct traffic as website visits that arrived on your site either by typing your website URL into a browser or through browser bookmarks. Organic search traffic is defined as website visits resulting from unpaid listings in search engine results. Social traffic is website visits through social media networks. Referral traffic is defined as website visits through other website domains. Google Analytics uses “unique identifiers” through the use of cookies to associate website visits with a particular user [23].

Data were compiled to assess the level of outreach. By monitoring both viewership and engagement, 2 layers of understanding could be gathered. The first aim was to determine the success of views among the target demographic. The second aim focused on how many of those users found the content interesting enough to engage or sign up for. By analyzing both of these variables, points of improvement were identified and adjusted to understand the full potential of humor-based advertising.

Based on the successful increase in social media engagement, which subsequently resulted in more frequent website traffic, according to the laugh model [19], an analysis of visitors to the GetWaivered website was compared to previous data before the implementation of humor-based advertisements was explored. The data was compiled to analyze daily visitors and the number of views per post. The demographics were further broken down into return users and new users. By assessing the chronology of posts and viewer retention, the team was able to assess the trends and growth of outreach at the end of the research period.

Results
From April 2021 to July 2021, there were a total of 9598 visits to the GetWaivered website. There was an average of 79.3 visitors per day, with the lowest number of daily visitors being 0 and the highest being 575 (Textbox 1).

The website page that was visited the most was the course registration page [22], which resulted in 5100 views and was where the e-training registration form, dates, and instructions were located.

Table 2 shows the differences between user engagement metrics (page views, users, and sessions) for the GW website in the months preceding the meme campaign (January to March 2021) and the meme campaign (April to July 2021), in addition to their respective P values. During the campaign dated April 2021 through July 2021, the statistical analysis has reflected an increase of 7602 in page views (P = .009), an increase of 5878 new users (P = .002), and an increase in website sessions of 5990 (P = .003).

The page received its most daily website traffic (575, as previously noted) on May 21, 2021, after the social media post, as indicated in Figure 3, was disseminated. After the creation and dissemination of Figure 4 and the website receiving its most daily visitors, the month of May 2021 resulted in our greatest number of e-training registrants (243).

Textbox 1. Website traffic analytics.

<table>
<thead>
<tr>
<th>Website visits from April 2021 to July 2021</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Total visits: 9598</td>
</tr>
<tr>
<td>• Daily average: 79.3</td>
</tr>
<tr>
<td>• Daily high: 575</td>
</tr>
<tr>
<td>• Daily low: 0</td>
</tr>
</tbody>
</table>

https://humanfactors.jmir.org/2023/1/e40244

Jmir Hum Factors 2023 | vol. 10 | e40244 | p.196
Table 2. Timewise comparison of GetWaivered user engagement metrics.

<table>
<thead>
<tr>
<th></th>
<th>Page views</th>
<th>Users</th>
<th>Sessions</th>
</tr>
</thead>
<tbody>
<tr>
<td>January-March 2021, n</td>
<td>6853</td>
<td>4287</td>
<td>4562</td>
</tr>
<tr>
<td>April-July 2021, n</td>
<td>14,455</td>
<td>10,165</td>
<td>10,552</td>
</tr>
<tr>
<td>Difference (P value)</td>
<td>7602 (.009)</td>
<td>5878 (.002)</td>
<td>5990 (.003)</td>
</tr>
</tbody>
</table>

Figure 3. GetWaivered pageviews, users, and sessions from January 2021 through July 2021. A timewise display of user engagement metrics (A: pageviews; B: users; C: sessions) shows substantial increases in metrics from January 2021 to March 2021 versus April 2021 to July 2021.

Figure 4. Memes correlated with GW’s largest e-course. GW: GetWaivered.

Discussion

Principal Findings

Our findings are key to informing future GW social media strategies and also provide an outline for other public health social media accounts, especially those targeting medical education. Since methods for analyzing social media use in public health are varied, these preliminary findings provide a basis for developing social media posting structures and evaluating them for future projects.

Website analytics were gathered prior to the July 22, 2021, e-training to assess the website traffic. The discontinuation of the data analytics was due to the removal of promotional means (eg, memes) before the scheduled July training and allowed for analytics and comparisons of training, both with and without promotion. Data on other population indicators, such as gender, age group, and location of followers on social media, were also gathered but not specifically analyzed during the time of this study.

The target audience was able to comment, react, and especially share information after viewing the memes. The opportunity for the viewer to share the post would provide a wider network of people to have access to the training.

Another important feature of our study was that the targeted population had the opportunity to register for training immediately after viewing the memes (Figure 4). This option for immediate action by providing the registration link in the description of the posts would remove any questions about what his or her next steps may look like. The individual was not only
allowed to share the information but to also take a personal step in his or her career to incite change in the local community.

Conclusions
The incorporation of the laugh model framework assisted in the promotion of GW’s remote digital DEA-X waiver training. The correlation between our pre- and postmeme promotional metrics showed the potential for our meme-based strategy as a method for community health intervention.

This finding shows the potential for humor-based promotional methods to show effectiveness in community engagement when compared to previous methods of educationally based promotions. This adds to the evidence that digitally native approaches can be a large driver in the future of promotional approaches that GW will use to enhance and increase course registrations and may be applied to other scenarios where similar communication is required.

Limitations
One limitation of this study is that the data used for analysis is for 121 days. Future studies will focus on specific meme campaigns that span longer periods and throughout more GW courses. Furthermore, we have not yet deployed website registration tracking from the links in the social media posts, but we were able to track total course registrants in real-time. However, we aim to have the tracking implemented in future courses to be able to add clarity to the relationship between social media memes and actual course awareness. Memes and comedic posts are only one form of engaging social media content. There are a variety of other engaging posts that can be used, such as GIFs, article retweets, incentive advertisements, positive messages or stories, and influencer posts. We suggest further research to investigate these other avenues and compare memes to other forms of engaging posts as well.

Comparison With Prior Work
Social media can and has been a useful tool for multiple public health efforts, as it allows not only for users to engage and interact with one another on social media platforms but can also allow for marketers, such as public health organizations, to engage with the users as well [24]. Public health organizations have been steadily improving their digital presence, but several of these social media platforms are not impactful due to a lack of user engagement [19]. Various studies have been done on the different uses of social media by using both memes and social media influencers to impact change. One study done in relation to the Truth Initiative measured the interactions from tweets and memes by social media influencers to discourage the use of tobacco products [25]. Memes, such as “CATmeggon,” were made to have a positive and comedic tone and were found to reach close to 1.5 million people per day [25].

As social media users are more likely to use social media platforms as means for passing their time or simply as means for instant gratification, one study recommends that public health organizations use the laugh model in order to engage with users [19]. The laugh model that was recommended in the study incorporates factors such as humor, viral content, and entertainment to effectively promote information to the public [19]. Social media campaigns have been used to increase awareness of the COVID-19 pandemic [12]. One study reported that social media was an essential tool in promoting behavioral changes and ways to increase protection against the novel virus [12]. Various studies have been done on the different uses of social media, using both memes and social media influencers to impact change [25]. One study done in relation to the Truth Initiative measured the interactions from tweets and memes by social media influencers to discourage the use of tobacco products [25]. Another study by Brown et al [26] incorporated memes into pharmacy education. The study revealed that the incorporation of memes, though impactful in engaging with the current-age audience, needs more research to be done to make any conclusions on their effectiveness [26]. Our study shows a substantial increase in the number of clinicians who were able to attend the DEA-X waiver training by simply framing public health issues using the methods discussed above.

Future Work
The future work of GW focuses on expanding social media campaigns and outreach to increase viewership and engagement. Our GW team plans to deploy more engaging content posts and memes and increase promotional spending on Facebook to reach more people. We can also focus on action plans for long-term engagement with users and collecting data from future courses for prospective studies. Furthermore, we aim to investigate the ideal combination of social media platform strategies for course enrollment and the impact of additional social media platforms such as LinkedIn and YouTube. This evaluation provides preliminary data and a framework as a baseline for planned future studies.

Acknowledgments
This study was funded by the Foundation for Opioid Response Efforts (FORE). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies or stance, either expressed or implied, of FORE. FORE is authorized to reproduce and distribute reprints for Foundation purposes, notwithstanding any copyright notation hereon.

The authors would also like to acknowledge Substance Abuse and Mental Health Services Administration and the California Bridge Program for their continued support of GW and our efforts to address the opioid epidemic. We would also like to extend our appreciation to Paul Faucher for his advisory services.
Conflicts of Interest

SH is employed by the Mass General Physician Organization and Mass General Institute of Health Professions; is part of the Advisory Board of Covid Act Now and Safer.App; is a cofounder of Executive Board ConductScience Inc; is part of the American College of Emergency Physician Supply Chain Task Force; has received research funding from the Foundation for Opioid Response Efforts (FORE) and personal fees from Withings Inc, Boston Globe, American College of Emergency Physicians, MazeEngineers (Maze Eng Inc), ConductScience Inc, Curative Medical Associates, VIO Med Spa New England, and Sci Sprout LLC; and is a volunteer at Emojination.

References

1. Buprenorphine. SAMHSA: Substance Abuse and Mental Health Service Administration. URL: https://www.samhsa.gov/medication-assisted-treatment/medications-counseling-related-conditions/buprenorphine [accessed 2021-02-20]


6. About Twitter | Our company and priorities. Twitter. URL: https://about.twitter.com/ [accessed 2021-02-20]


20. Utah family meals. Tooele County Health Department. URL: https://tooelehealth.org/utah-family-meals/ [accessed 2023-01-18]


23. How users are identified for user metrics. Google Analytics Help. URL: https://support.google.com/analytics/answer/2992042?hl=en [accessed 2022-03-17]


Abbreviations

- DEA: Drug Enforcement Administration
- FORE: Foundation for Opioid Response Efforts
- GW: GetWaivered
- NP: nurse practitioner
- OUD: opioid use disorder
- PA: physician assistant

©Darrel Wang, Neha Balapal, Amala Ankem, Saishravan Shyamsundar, Adarsh Balaji, Jasmine Kannikal, Marlie Bruno, Shuhan He, Paul Chong. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 27.01.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Persuasive Messages for Improving Adherence to COVID-19 Prevention Behaviors: Randomized Online Experiment

Mehdi Mourali¹, PhD; Jamie L Benham², MD, PhD; Raynell Lang³, MD, PhD; Madison M Fullerton³, MSc; Jean-Christophe Boucher⁴, PhD; Kirsten Cornelson⁵, PhD; Robert J Oxoby⁶, PhD; Cora Constantinescu⁷, MD, PhD; Theresa Tang⁷, BHSc; Deborah A Marshall²,³, PhD; Jia Hu³, MD

¹Haskayne School of Business, University of Calgary, Calgary, AB, Canada
²Department of Medicine, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada
³Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada
⁴Department of Political Science, School of Public Policy, University of Calgary, Calgary, AB, Canada
⁵Department of Economics, University of Notre Dame, Notre Dame, IN, United States
⁶Department of Economics, Faculty of Arts, University of Calgary, Calgary, AB, Canada
⁷Department of Pediatrics, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

Corresponding Author:
Mehdi Mourali, PhD
Haskayne School of Business
University of Calgary
2500 University Drive NW
Calgary, AB, T2N 1N4
Canada
Phone: 1 403 220 6684
Email: mehdi.mourali@haskayne.ucalgary.ca

Abstract

Background: Adherence to nonpharmaceutical interventions for COVID-19, including physical distancing, masking, staying home while sick, and avoiding crowded indoor spaces, remains critical for limiting the spread of COVID-19.

Objective: The aim of this study was to test the effectiveness of using various persuasive appeals (deontological moral frame, empathy, identifiable victim, goal proximity, and reciprocity) at improving intentions to adhere to prevention behaviors.

Methods: A randomized online experiment using a representative sample of adult Canadian residents with respect to age, ethnicity, and province of residence was performed from March 3 to March 6, 2021. Participants indicated their intentions to follow public health guidelines, saw one of six flyers featuring a persuasive appeal or no appeal, and then rated their intentions a second time. Known correlates of attitudes toward public health measures were also measured.

Results: Intentions to adhere to public health measures increased in all appeal conditions. The message featuring an empathy appeal resulted in a greater increase in intentions than the control (no appeal) message. Moreover, the effectiveness of persuasive appeals was moderated by baseline intentions. Deontological, empathy, identifiable victim, and reciprocity appeals improved intentions more than the control message, but only for people with lower baseline intentions to adhere to nonpharmaceutical interventions.

Conclusions: Public health marketing campaigns aiming to increase adherence to COVID-19 protective behaviors could achieve modest gains by employing a range of persuasive appeals. However, to maximize impact, it is important that these campaigns be targeted to the right individuals.

Trial Registration: ClinicalTrials.gov NCT05722106; https://clinicaltrials.gov/ct2/show/NCT05722106

(JMIR Hum Factors 2023;10:e41328) doi:10.2196/41328

KEYWORDS
COVID-19; messaging; persuasion; behavioral intentions; public health; compliance; prevention; physical distance; mask; sick; effectiveness; behavior
Introduction

Background
As of July 2022, over 500 million people worldwide have contracted the SARS-CoV-2 virus, resulting in over 6 million COVID-19–related deaths [1]. Despite the remarkable and ongoing effort to inoculate the world population (over 12 billion vaccine doses have been administered so far), the rapidly evolving virus continues to spread at alarmingly high rates. Even affluent countries like Canada—a G7 member with over 83% of the population fully vaccinated—are struggling to contain the spread, with case and hospitalization numbers reaching all-time highs in the winter of 2022 [2,3]. With governments gradually lifting restrictive measures and reopening borders, it is critical that, in addition to getting vaccinated, individuals continue to follow nonpharmaceutical interventions—including wearing face masks, physical distancing, staying home when ill, and avoiding crowded indoor spaces—to limit the spread of this highly transmissible virus, especially as newer more transmissible variants continue to emerge [4-7].

Mandates and government-imposed restrictions are important policy tools for limiting the spread of COVID-19, but they are insufficient on their own and must be complemented by softer interventions designed to increase compliance with public health guidelines. Convincing citizens to freely adhere to social distancing, masking, and other preventive behaviors requires persuasive communication going beyond providing information on the risks of the pandemic. Public health organizations and governments need to understand how to best frame messages to effectively appeal to different audiences [8].

The primary objective of this study was to empirically test the effectiveness of message framings emphasizing a set of carefully selected persuasive appeals at improving people’s intentions to engage in health protective behaviors. Another aim of the study was to characterize the target audience most susceptible to respond positively to the persuasive appeals. The findings are intended to guide the design and development of public health campaigns in Canada.

Message Framing and Adherence to Public Health Measures
In the past year, numerous studies have investigated the impact of various persuasive appeals on people’s attitudes and intentions around COVID-19–related behaviors. The studies varied in their methods and procedures and produced mixed results. Messages using prosocial, altruistic, other-focused, or community-focused appeals were generally more persuasive than messages using self-interested, self-protective, or threatening appeals [9-16]. Likewise, gain-framed messages were typically more effective than loss-framed messages [17,18], although at least one study found the opposite result [19]. Moreover, messages invoking social norms do not seem to be particularly effective [20,21].

In a comprehensive analysis, Pink and colleagues [21] tested 56 short messages using a wide range of framings, including some of the appeals mentioned above. They found no consistent effects for any of the tested messages. Nevertheless, a message using a reciprocity appeal performed the best in three of their five studies.

The present research adds to this body of work by testing the effectiveness of five appeals (deontological moral frame, empathy, goal proximity, identifiable victim, and reciprocity) at improving people’s intentions to adhere to public health measures. This study differs from prior work in at least two important aspects. First, the pandemic context at the time of our study (early March 2021) is unlike that characterizing the early stages of the pandemic when most previous studies were conducted. At the time of our study, there had been over 880,000 confirmed COVID-19 cases in Canada, including over 22,000 deaths. Vaccine supply was limited with just over 2 million doses administered by March 3, 2021 [2]. Although the daily COVID-19 activity had been declining from mid-January through mid-February, it has leveled off since. The 7-day average was under 3000 new cases a day nationwide, but variants of concern (B.1.1.7 and B1.351) had emerged [22]. Masking in public places was mandated in most jurisdictions, and the public was advised to limit travel and minimize contact with people outside of their household [22]. The difference in context alone may result in notable differences in how people process and respond to various persuasive messages.

Previous experiments have largely neglected the role of baseline attitudes and intentions when testing for differences between messages. In contrast, we expected baseline intentions to have a significant impact on how people respond to persuasive messages. People who are highly compliant to begin with have little room left for improvement. Thus, we expected the effect of persuasive appeals to be stronger among those with relatively lower baseline intentions. This is significant because those who are less compliant with public health measures are a critical target for behavior change.

Five Persuasive Appeals
This study focused on the impact of five persuasive appeals: deontological moral frame, empathy, identifiable victim, goal proximity, and reciprocity. Deontological moral frames are frequently encountered in the current public discourse; they appeal to the sense of duty and responsibilities we have to our families and communities [23]. Prior research suggests that agents making deontological judgments are perceived to be more trustworthy than agents making utilitarian judgments [24,25], even when they are not actually more trustworthy [26]. Moreover, research using machine learning found that moral identity is a strong predictor of adherence to public health measures [27]. Thus, we expect persuasive appeals that use deontological moral frames to help increase adherence to public health measures.

Empathy—understanding and feeling concerned for vulnerable others—has been found to increase altruism and caring, and to motivate helping behavior [28-30]. Thus, inducing empathy by highlighting that the sick, elderly, and immunocompromised need our help is expected to increase adoption of health protective behaviors [13,15].
Goal-proximity appeals emphasize that better days are approaching. This is important because people’s motivation to comply with public health advice has declined since the pandemic’s early days. A Gallup study tracking social distancing behaviors found that the percentage of Americans practicing social distancing dropped steadily over time, from 75% in April 2020 to 38% in March 2021 [31]. A drop in motivation over the course of goal pursuit is not uncommon when pursuing goals with no clear end states or when the tasks required to achieve the goal are difficult [32]. Fortunately, motivational strength tends to increase as the distance to the goal decreases. The goal-gradient hypothesis holds that people apply more effort and persistence as they get closer to a goal’s end state [33-37]. The third message tested in this study relies on this motivational property.

The fourth message relies on the persuasive power of identifiable victims. The identifiable victim effect refers to people’s propensity to offer more help to specific, identifiable victims rather than to anonymous, statistical victims [38-40]. This effect has been attributed to the fact that identifiable victims evoke more powerful emotional responses than statistical victims [38,41]. The identifiable victim effect also arises because people believe their contribution will have a greater impact on an identified victim than on a large group of unidentified victims [39].

Our fifth message relies on the principle of reciprocity. According to Cialdini [42], “all societies subscribe to a norm that obligates individuals to repay in kind what they have received” (page 76). The reciprocity code is not limited to gifts and favors but also includes concessions, whereby people are more likely to make concessions to those who have made concessions to them [43,44]. Accordingly, our reciprocity message emphasizes the sacrifices health care workers are making to help and protect us, and asks that we return the favor by adhering to health protective behaviors.

Individual Differences in Compliance With Public Health Measures

We expect persuasive communication to have a greater impact among individuals who have lower initial intentions to adhere with public health measures. This is because individuals who have high initial intentions have little room left for improvement; that is, they are already persuaded and further exposure to persuasive communication is unlikely to change their intentions. From a campaign planning perspective, it is important to identify who these individuals might be so that the messages can be efficiently targeted.

The existing literature points to significant variability in the levels of adherence to public health measures [45-52]. A recent review of 29 empirical studies concluded that greater adherence to public health measures is reliably associated with being older, identifying as female, trusting governments, perceiving COVID-19 as a threat, and accessing information through traditional news media [50]. Variability in uptake of public health behaviors was also linked to differences in political ideology [51,52] and perceived responsibility for others [53]. In this study, we measured these characteristics and examined their associations with baseline intentions.

Methods

Participants and Procedure

A representative sample of adult Canadian residents with respect to age, ethnicity, and province of residence was recruited by the research firm Critical Mass between March 3 and March 6, 2021. A description of the study was posted on Lucid Marketplace, a third-party platform that maintains an online research panel of 15 million verified users. Users from Canada were invited to visit a screening page assessing demographic and geographic variables. Target quotas for province of residence, age, gender, and ethnicity were set to obtain a demographically representative sample based on the 2016 census data (see Table S1 in Multimedia Appendix 1 for details on the quota system).

Upon consenting in writing, participants reported on their intentions to engage in a set of prevention behaviors over the coming weeks (T1). They were then randomly assigned to an active control or one of five persuasive appeal conditions (control vs deontological vs empathy vs goal proximity vs reciprocity vs identifiable victim) and reported on their intentions to engage in the same set of prevention behaviors a second time (T2). This design allowed us to examine whether the effectiveness of persuasive appeals varies as a function of initial prevention intentions. Finally, participants completed a series of questions assessing potential correlates of prevention intentions. These included measures of political orientation, trust in institutions, perceived threat of COVID-19, and perceived responsibility toward others.

Ethics Approval

This study was approved by the University of Calgary Conjoint Research Ethics Board (REB21-0173) and was conducted according to the principles expressed in the Declaration of Helsinki.

Measures

Index variables for intentions to engage in prevention behaviors (pre- and posttreatment) were created by averaging across six items: (1) Limit my physical contact with others when possible, (2) Completely avoid any unnecessary physical contact with others (eg, hugging or handshakes), (3) Avoid crowded indoor spaces, (4) Wear a mask when I leave the house, (5) Wash my hands as much as possible, and (6) Stay home when mildly sick. These items were measured on 100-point sliding scales (0=strongly disagree, 50=neither agree nor disagree, 100=strongly agree).

Persuasive appeals were manipulated using promotional flyers ostensibly distributed by the Public Health Agency of Canada. In the control condition, the flyer contained a simple list of what participants can do to help prevent the spread of COVID-19. In each of the five persuasion conditions, the flyer contained the same basic information and a unique persuasive appeal (see Figure 1 for an example and Figures S1-S5 in Multimedia Appendix 1 for the remaining flyers). The wording of the messages is shown in Textbox 1.
Trust in various institutions (politicians, civil servants, public health officials, physicians, other health care providers [e.g., nurses, pharmacists], scientists, journalists, and pharmaceutical companies) was measured using eight items (α=.91) on 100-point sliding scales (0=do not trust at all, 100=trust completely).

Perceived COVID-19 threat was measured using four items (α=.89) adapted from previous research [11]. A sample item is: “To what extent are you afraid of contracting COVID-19 because of the consequences for you personally/your community?” (0=not at all, 50=to a moderate extent, 100=to an enormous extent).

Perceived responsibility toward others was assessed using four items (α=.94) adapted from previous research [18]. A sample item is: “I owe it to my family to do whatever I can to stop the spread of COVID-19” (1=strongly disagree, 7=strongly agree).

Finally, political orientation was measured using the following item: “If you think about your own political views, where would you classify your views on this scale?” (1=very liberal, 7=very conservative).

Figure 1. Sample flyer: empathy appeal.
Textbox 1. Messages across appeal conditions.

**Control**
The virus spreads mainly between people who are in close contact with one another. You can help prevent the spread of COVID-19. We can all do our part:

- Avoid social gatherings.
- Wear a mask when you go out.
- Stay at least six feet away from people outside your household.
- Wash your hands often with soap and water.

These actions prevent the spread of COVID-19.

**Deontological**
The virus spreads mainly between people who are in close contact with one another. You can help prevent the spread of COVID-19. We can all do our part:

- Avoid social gatherings.
- Wear a mask when you go out.
- Stay at least six feet away from people outside your household.
- Wash your hands often with soap and water.

We all need to do this, however difficult, because it is the right thing to do: it is our duty and responsibility to protect our families, friends, and fellow citizens.

**Empathy**
The sick, elderly, and immunocompromised need our help. We all have a choice. If we don’t take the right actions, we risk the lives of others. But we can protect those most likely to be harmed. We can protect those who are vulnerable by taking simple steps:

- Avoid social gatherings.
- Wear a mask when you go out.
- Stay at least six feet away from people outside your household.
- Wash your hands often with soap and water.

Take action to protect those who are vulnerable!

**Identifiable victim**
A few weeks ago, Sam was a healthy 26-year-old with no medical complications. Then he suddenly came down with a bad cough and a feeling like he could not breathe. He tested positive for COVID-19 and is now hospitalized, receiving oxygen from a ventilator, and fighting for his life. This could be any of us. Reduce the risk to yourself and others:

- Avoid social gatherings.
- Wear a mask when you go out.
- Stay at least six feet away from people outside your household.
- Wash your hands often with soap and water.

If we take these actions, we can prevent more people from suffering the way Sam has.

**Goal proximity**
The recent development of safe and effective vaccines gives us great hope. We see the light at the end of the tunnel, but we are not quite there yet. Until a large proportion of the population is immunized, we must remain vigilant and double our efforts to prevent the spread of COVID-19.

- Avoid social gatherings.
- Wear a mask when you go out.
- Stay at least six feet away from people outside your household.
- Wash your hands often with soap and water.

These actions prevent the spread of COVID-19.

**Reciprocity**
Doctors, nurses, and other health care workers are working around the clock, often risking their lives to care for patients with the coronavirus. Working long hours in highly infectious environments, many of them are falling ill. As our health care workers put their lives on the line, we can do our part:

- Avoid social gatherings.
- Wear a mask when you go out.
- Stay at least six feet away from people outside your household.
- Wash your hands often with soap and water.

Our brave health care workers have sacrificed to help others. We should take action too.

Data Analysis

First, we sought to address the broad question: does exposure to messages using persuasive appeals improve intentions to engage in prevention behaviors more than exposure to the control message? Given the structure in our data (each participant provided two sets of ratings), we fitted a linear mixed effects model (estimated using maximum likelihood) with intention to engage in prevention behaviors as the outcome variable; random intercepts for participants (id); and fixed effects for appeal condition, time of rating, and their interaction. In this analysis, the \( P \) values were estimated via \( t \)-tests using the Satterthwaite approximation to degrees of freedom. Effect sizes for the fixed effects are indicated by the standardized regression coefficients (\( \beta \)) and their 95% CIs.

We performed a series of moderated regressions (estimated using ordinary least squares [OLS]) to investigate whether the effectiveness of persuasive appeals varies as a function of baseline prevention intentions. We used change in intentions as the outcome variable, persuasion appeal as a binary predictor, and baseline intentions as a continuous moderator.

To help characterize the target audience, we examined the association of baseline intentions with demographic variables, including age, gender, ethnic background, education, and geographic region, as well as attitudinal variables such as perceived COVID-19 threat, perceived responsibility toward others, trust in institutions, and political orientation.

We fitted a linear model (estimated using OLS) using all predictors. The continuous predictors (age, threat, responsibility, trust, and political orientation) were mean-centered and the categorical predictors were dummy-coded. The ethnic background variable was constructed by recoding the original ethnicity variable into a binary variable (0=ethnic majority, 1=ethnic minority). Education was modified by combining the “less than high school” and “high school” categories into a single “high school or less” category, which served as the baseline group in the analysis. The region variable was constructed by collapsing the Newfoundland and Labrador, Nova Scotia, New Brunswick, and Territories categories in the province variable into a single “Maritimes and Territories” category. Ontario was set as the baseline category for the five-level region variable and female was set as the baseline category for the three-level gender variable.

Data analysis was performed using the statistical program R version 4.0.2 [54], and the level of statistical significance was set at \( \alpha=0.05 \).

Results

Participant Characteristics

A total of 7079 respondents visited the screening page. Of those, 3746 qualified for the main study based on the quota requirements. Of the qualified respondents, 78 failed to complete the survey, resulting in a final sample of 3668 participants (see Table 1 for sample characteristics). Those who failed to complete the survey were demographically similar to those who completed the survey, but were predominantly from the provinces of Quebec (40%) and Nova Scotia (19%) (see Table S1 in Multimedia Appendix 1).
<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Overall (N=3668), n (%)</th>
<th>Control (n=582), n (%)</th>
<th>Deontological (n=622), n (%)</th>
<th>Empathy (n=624), n (%)</th>
<th>Proximity (n=603), n (%)</th>
<th>Reciprocity (n=623), n (%)</th>
<th>Victim (n=614), n (%)</th>
<th>P value&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong> (n=3668)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.27</td>
</tr>
<tr>
<td>Female</td>
<td>2202 (60.03)</td>
<td>334 (57.4)</td>
<td>380 (61.1)</td>
<td>353 (56.6)</td>
<td>359 (59.5)</td>
<td>386 (62.0)</td>
<td>390 (63.5)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1450 (39.53)</td>
<td>245 (42.1)</td>
<td>238 (38.3)</td>
<td>267 (42.8)</td>
<td>243 (40.3)</td>
<td>234 (37.6)</td>
<td>223 (36.3)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>16 (0.44)</td>
<td>3 (0.5)</td>
<td>4 (0.6)</td>
<td>4 (0.6)</td>
<td>1 (0.2)</td>
<td>3 (0.5)</td>
<td>1 (0.2)</td>
<td></td>
</tr>
<tr>
<td><strong>Age group (years)</strong> (n=3667)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.73</td>
</tr>
<tr>
<td>18-24</td>
<td>345 (9.41)</td>
<td>54 (9.3)</td>
<td>65 (10.5)</td>
<td>60 (9.6)</td>
<td>52 (8.6)</td>
<td>54 (8.7)</td>
<td>60 (9.8)</td>
<td></td>
</tr>
<tr>
<td>25-34</td>
<td>690 (18.82)</td>
<td>118 (20.3)</td>
<td>118 (19.0)</td>
<td>125 (20.0)</td>
<td>115 (19.1)</td>
<td>113 (18.1)</td>
<td>101 (16.4)</td>
<td></td>
</tr>
<tr>
<td>35-44</td>
<td>785 (21.41)</td>
<td>119 (20.4)</td>
<td>125 (20.1)</td>
<td>146 (23.4)</td>
<td>145 (24.0)</td>
<td>131 (21.0)</td>
<td>119 (19.4)</td>
<td></td>
</tr>
<tr>
<td>45-54</td>
<td>599 (16.33)</td>
<td>97 (16.7)</td>
<td>106 (17.0)</td>
<td>101 (16.2)</td>
<td>86 (14.3)</td>
<td>103 (16.5)</td>
<td>106 (17.3)</td>
<td></td>
</tr>
<tr>
<td>55-64</td>
<td>595 (16.23)</td>
<td>100 (17.2)</td>
<td>103 (16.6)</td>
<td>91 (14.6)</td>
<td>102 (16.9)</td>
<td>100 (16.1)</td>
<td>99 (16.1)</td>
<td></td>
</tr>
<tr>
<td>65-99</td>
<td>653 (17.81)</td>
<td>94 (16.2)</td>
<td>105 (16.9)</td>
<td>101 (16.2)</td>
<td>103 (17.1)</td>
<td>121 (19.4)</td>
<td>129 (21.0)</td>
<td></td>
</tr>
<tr>
<td><strong>Ethnicity</strong> (n=3650)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.28</td>
</tr>
<tr>
<td>White</td>
<td>2840 (77.81)</td>
<td>448 (77.0)</td>
<td>479 (77.0)</td>
<td>478 (76.6)</td>
<td>465 (77.1)</td>
<td>488 (78.3)</td>
<td>482 (78.5)</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>110 (3.01)</td>
<td>21 (3.6)</td>
<td>21 (3.4)</td>
<td>18 (2.9)</td>
<td>16 (2.7)</td>
<td>15 (2.4)</td>
<td>19 (3.1)</td>
<td></td>
</tr>
<tr>
<td>East Asian</td>
<td>297 (8.14)</td>
<td>48 (8.3)</td>
<td>49 (7.9)</td>
<td>63 (10.1)</td>
<td>44 (7.3)</td>
<td>47 (7.6)</td>
<td>46 (7.6)</td>
<td></td>
</tr>
<tr>
<td>South Asian</td>
<td>193 (5.29)</td>
<td>29 (5.0)</td>
<td>27 (4.4)</td>
<td>34 (5.5)</td>
<td>30 (5.0)</td>
<td>42 (6.8)</td>
<td>31 (5.1)</td>
<td></td>
</tr>
<tr>
<td>Indigenous</td>
<td>63 (1.73)</td>
<td>12 (2.1)</td>
<td>13 (2.1)</td>
<td>10 (1.6)</td>
<td>6 (1.0)</td>
<td>11 (1.8)</td>
<td>11 (1.8)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>147 (4.03)</td>
<td>21 (3.6)</td>
<td>29 (4.7)</td>
<td>18 (2.9)</td>
<td>40 (6.7)</td>
<td>19 (3.1)</td>
<td>20 (3.3)</td>
<td></td>
</tr>
<tr>
<td><strong>Education</strong> (n=3667)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.23</td>
</tr>
<tr>
<td>Less than high school</td>
<td>86 (2.35)</td>
<td>12 (2.1)</td>
<td>11 (1.8)</td>
<td>12 (1.9)</td>
<td>14 (2.3)</td>
<td>26 (4.2)</td>
<td>11 (1.8)</td>
<td></td>
</tr>
<tr>
<td>High school</td>
<td>718 (19.58)</td>
<td>112 (19.2)</td>
<td>108 (17.4)</td>
<td>127 (20%)</td>
<td>107 (17.7)</td>
<td>121 (19.4)</td>
<td>143 (23.3)</td>
<td></td>
</tr>
<tr>
<td>Some college</td>
<td>631 (17.21)</td>
<td>98 (16.8)</td>
<td>111 (17.8)</td>
<td>101 (16%)</td>
<td>113 (18.7)</td>
<td>100 (16.1)</td>
<td>108 (17.6)</td>
<td></td>
</tr>
<tr>
<td>College</td>
<td>834 (22.74)</td>
<td>128 (22.0)</td>
<td>155 (24.9)</td>
<td>125 (20.4)</td>
<td>143 (23.7)</td>
<td>149 (23.9)</td>
<td>134 (21.8)</td>
<td></td>
</tr>
<tr>
<td>University</td>
<td>1007 (27.46)</td>
<td>169 (29.0)</td>
<td>174 (28.0)</td>
<td>185 (29.6)</td>
<td>156 (25.9)</td>
<td>165 (26.5)</td>
<td>158 (25.7)</td>
<td></td>
</tr>
<tr>
<td>Graduate degree</td>
<td>391 (10.66)</td>
<td>63 (10.8)</td>
<td>63 (10.1)</td>
<td>74 (11.9)</td>
<td>69 (11.4)</td>
<td>62 (10.0)</td>
<td>60 (9.8)</td>
<td></td>
</tr>
<tr>
<td><strong>Province</strong> (n=3668)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.79</td>
</tr>
<tr>
<td>Newfoundland and Labrador</td>
<td>74 (2.02)</td>
<td>6 (1.0)</td>
<td>18 (2.9)</td>
<td>13 (2.1)</td>
<td>9 (1.5)</td>
<td>14 (2.2)</td>
<td>14 (2.3)</td>
<td></td>
</tr>
<tr>
<td>Prince Edward Island</td>
<td>19 (0.52)</td>
<td>3 (0.5)</td>
<td>5 (0.8)</td>
<td>4 (0.6)</td>
<td>4 (0.7)</td>
<td>2 (0.3)</td>
<td>1 (0.2)</td>
<td></td>
</tr>
<tr>
<td>New Brunswick</td>
<td>96 (2.62)</td>
<td>11 (1.9)</td>
<td>20 (3.2)</td>
<td>21 (3.4)</td>
<td>15 (2.5)</td>
<td>10 (1.6)</td>
<td>19 (3.1)</td>
<td></td>
</tr>
<tr>
<td>Nova Scotia</td>
<td>122 (3.33)</td>
<td>22 (3.8)</td>
<td>23 (3.7)</td>
<td>24 (3.8)</td>
<td>16 (2.7)</td>
<td>22 (3.5)</td>
<td>15 (2.4)</td>
<td></td>
</tr>
<tr>
<td>Quebec</td>
<td>472 (12.87)</td>
<td>72 (12.4)</td>
<td>87 (14.0)</td>
<td>83 (13.3)</td>
<td>76 (12.6)</td>
<td>66 (10.6)</td>
<td>88 (14)</td>
<td></td>
</tr>
<tr>
<td>Ontario</td>
<td>1555 (42.39)</td>
<td>256 (44.0)</td>
<td>259 (41.6)</td>
<td>251 (40.2)</td>
<td>267 (44.3)</td>
<td>273 (43.8)</td>
<td>249 (40.6)</td>
<td></td>
</tr>
<tr>
<td>Manitoba</td>
<td>155 (4.23)</td>
<td>26 (4.5)</td>
<td>26 (4.2)</td>
<td>26 (4.2)</td>
<td>26 (4.3)</td>
<td>30 (4.8)</td>
<td>21 (3.4)</td>
<td></td>
</tr>
<tr>
<td>Saskatchewan</td>
<td>128 (3.49)</td>
<td>18 (3.1)</td>
<td>21 (3.4)</td>
<td>22 (3.5)</td>
<td>18 (3.0)</td>
<td>20 (3.2)</td>
<td>29 (4.7)</td>
<td></td>
</tr>
<tr>
<td>Alberta</td>
<td>470 (12.81)</td>
<td>76 (13.1)</td>
<td>77 (12.4)</td>
<td>84 (13.5)</td>
<td>65 (10.8)</td>
<td>86 (13.8)</td>
<td>82 (13.4)</td>
<td></td>
</tr>
<tr>
<td>British Columbia</td>
<td>569 (15.51)</td>
<td>89 (15.3)</td>
<td>85 (13.7)</td>
<td>95 (15.2)</td>
<td>105 (17.4)</td>
<td>100 (16.1)</td>
<td>95 (15.5)</td>
<td></td>
</tr>
<tr>
<td>Territories&lt;sup&gt;b&lt;/sup&gt;</td>
<td>8 (0.22)</td>
<td>3 (0.5)</td>
<td>1 (0.2)</td>
<td>1 (0.2)</td>
<td>2 (0.3)</td>
<td>0 (0)</td>
<td>1 (0.2)</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Pearson $\chi^2$ test.

<sup>b</sup>Territories=Yukon, Northwest Territories, and Nunavut.

---

https://humanfactors.jmir.org/2023/1/e41328

JMIR Hum Factors 2023 | vol. 10 | e41328 | p.207

(page number not for citation purposes)
Intentions to Engage in Prevention Behaviors

The results of the fixed factors in the mixed effects model are summarized in Table 2 (random effects: $\sigma^2=18.90$, $\tau_{00id}=282.54$, intraclass correlation coefficient=0.94, $N_{id}=3668$, observations=7331, marginal $R^2=0.006$, conditional $R^2=0.938$). Prior to exposure to the persuasive appeals, participants in all conditions reported similarly high intentions to engage in prevention behaviors. Prevention scores at T1 did not differ significantly between any appeal condition and the control condition, as shown in Table 2 (P values for deontological, empathy, goal proximity, reciprocity, and victim are all greater than .05). This confirmed that random assignment produced groups with equivalent baselines. Furthermore, exposure to a reminder message about prevention behaviors (ie, control condition) increased participants’ intentions to engage in prevention behaviors (see Time [T2] variable in Table 2). Additionally, exposure to a persuasive message using an empathy appeal resulted in a larger increase in intentions to engage in prevention behaviors relative to the control message (Table 2).

Table 3. Intention to engage in prevention behaviors before (T1) and after (T2) exposure to various appeals.

<table>
<thead>
<tr>
<th>Appeal</th>
<th>Intention_T1</th>
<th>Intention_T2</th>
<th>T2–T1</th>
<th>$t$ statistic</th>
<th>df</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>87.1</td>
<td>89.2</td>
<td>2.1</td>
<td>8.83</td>
<td>581</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Deontological</td>
<td>87.5</td>
<td>90.1</td>
<td>2.6</td>
<td>10.83</td>
<td>620</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Empathy</td>
<td>86.5</td>
<td>89.7</td>
<td>3.2</td>
<td>11.73</td>
<td>622</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Proximity</td>
<td>86.6</td>
<td>88.8</td>
<td>2.2</td>
<td>8.71</td>
<td>600</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Reciprocity</td>
<td>87.8</td>
<td>90.5</td>
<td>2.7</td>
<td>11.86</td>
<td>621</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Victim</td>
<td>87.5</td>
<td>90.2</td>
<td>2.7</td>
<td>10.10</td>
<td>613</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

T2: posttest point.
Moderating Effect of Baseline Intentions

The preceding analysis suggested that, apart from empathy, the use of persuasive appeals does not improve intentions to engage in prevention behaviors beyond a simple reminder message. However, we expected the effectiveness of persuasive appeals to vary according to people’s initial dispositions. Persuasive appeals are likely effective when baseline intentions are relatively low, but may have a limited impact when baseline intentions are so high that there is little room for improvement. Results from the moderated regressions were consistent with our expectations (see Table 4). The appeal × baseline intentions interaction was statistically significant for all but the goal-proximity appeal, suggesting that the effectiveness of the deontological, empathy, reciprocity, and identifiable victim appeals indeed depends on the level of initial intentions.

We followed up with floodlight analyses [55] of each significant interaction. As shown in Figure 3, the conditional effect of seeing a deontological appeal was significant only among participants who had a score of 85.5 or below on the initial intentions measure (30.2% of participants; mean 66.4). In other words, people with lower baseline intentions increased their intentions to engage in prevention behaviors more after seeing a message featuring a deontological appeal than after seeing a message featuring a simple reminder. In contrast, those with high baseline intentions (higher than 85.5; 69.8% of participants; mean 96.2) did not differ significantly in how much they changed their intentions when they saw a message featuring a deontological appeal or a message featuring a reminder.

We observed similar patterns with the other appeals. The conditional effect of empathy was significant only among participants scoring 90.1 or lower on initial intentions (39.5% of participants; mean 71.5), the conditional effect of reciprocity was significant only for those scoring 87.8 or lower on initial intentions (44.1% of participants; mean 68.7), and the conditional effect of identifiable victim was only significant for those scoring 84.8% or lower on initial intentions (29.3% of participants; mean 65.7).

Table 4. Effect of appeal × baseline intentions interaction on change in intentions to engage in prevention behavior.

<table>
<thead>
<tr>
<th>Appeal × baseline intentions</th>
<th>Estimate, b (SE)</th>
<th>t statistic</th>
<th>df</th>
<th>P value</th>
<th>β (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deontological</td>
<td>−0.08 (0.02)</td>
<td>−4.29</td>
<td>1199</td>
<td>&lt;.001</td>
<td>−0.12 (−0.17 to −0.06)</td>
</tr>
<tr>
<td>Empathy</td>
<td>−0.09 (0.02)</td>
<td>−4.60</td>
<td>1201</td>
<td>&lt;.001</td>
<td>−0.13 (−0.18 to −0.07)</td>
</tr>
<tr>
<td>Proximity</td>
<td>−0.02 (0.02)</td>
<td>−1.14</td>
<td>1179</td>
<td>.26</td>
<td>−0.03 (−0.09 to −0.02)</td>
</tr>
<tr>
<td>Reciprocity</td>
<td>−0.08 (0.02)</td>
<td>−4.38</td>
<td>1200</td>
<td>&lt;.001</td>
<td>−0.12 (−0.18 to −0.07)</td>
</tr>
<tr>
<td>Victim</td>
<td>−0.05 (0.02)</td>
<td>−2.75</td>
<td>1192</td>
<td>.006</td>
<td>−0.08 (−0.13 to −0.02)</td>
</tr>
</tbody>
</table>
Predictors of Baseline Intentions

The moderation analysis implied that a public health campaign using persuasive appeals would be most effective when targeting individuals with lower baseline intentions: but who might these individuals be?

The regression model using all demographic and attitudinal predictors explained a statistically significant and substantial proportion of the variance ($R^2=0.51$, $F_{16, 3415}=224.2$, $P<.001$, adjusted $R^2=0.51$). As shown in Table 5, baseline intentions increased with age, perception of COVID-19 threat, perceived responsibility, and trust in institutions. Conversely, baseline intentions decreased with political conservatism, were lower for males relative to females, and were lower in the Prairies compared to Ontario. Neither education level nor ethnic background was significantly uniquely associated with baseline intentions to engage in prevention behaviors.

Figure 3. Floodlight analysis of the interactive effects of appeal and baseline intentions. n.s: not significant ($P>.05$).
Table 5. Multivariable regression model of initial intentions.

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Estimate, b (SE)</th>
<th>t (df=3415)</th>
<th>P value</th>
<th>β (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>88.44 (0.57)</td>
<td>155.99</td>
<td>&lt;.001</td>
<td>.08 (.10 to .14)</td>
</tr>
<tr>
<td>Age</td>
<td>0.06 (0.01)</td>
<td>4.41</td>
<td>&lt;.001</td>
<td>.06 (.03 to .08)</td>
</tr>
<tr>
<td>Gender [Male]</td>
<td>−1.74 (0.44)</td>
<td>−3.94</td>
<td>&lt;.001</td>
<td>−10 (−.15 to −.05)</td>
</tr>
<tr>
<td>Gender [Other]</td>
<td>−3.01 (3.19)</td>
<td>−0.95</td>
<td>.34</td>
<td>−17 (−.53 to .18)</td>
</tr>
<tr>
<td>Ethnic [Minority]</td>
<td>−0.07 (0.54)</td>
<td>−0.13</td>
<td>.90</td>
<td>−0.06 (−.06 to .06)</td>
</tr>
<tr>
<td>Education [Some college]</td>
<td>−0.23 (0.68)</td>
<td>−0.35</td>
<td>.73</td>
<td>−0.1 (−.09 to .06)</td>
</tr>
<tr>
<td>Education [College]</td>
<td>−0.40 (0.63)</td>
<td>−0.63</td>
<td>.53</td>
<td>−0.09 (−.09 to .05)</td>
</tr>
<tr>
<td>Education [University]</td>
<td>0.38 (0.61)</td>
<td>0.63</td>
<td>.53</td>
<td>0.02 (−.05 to .09)</td>
</tr>
<tr>
<td>Education [Graduate degree]</td>
<td>−0.24 (0.79)</td>
<td>−0.30</td>
<td>.76</td>
<td>−0.1 (−.10 to .07)</td>
</tr>
<tr>
<td>Region [Maritimes]</td>
<td>−1.16 (0.79)</td>
<td>−1.46</td>
<td>.14</td>
<td>−0.07 (−.16 to .02)</td>
</tr>
<tr>
<td>Region [Quebec]</td>
<td>−0.22 (0.67)</td>
<td>−0.33</td>
<td>.74</td>
<td>−0.09 (−.09 to .06)</td>
</tr>
<tr>
<td>Region [Prairies]</td>
<td>−1.53 (0.57)</td>
<td>−2.68</td>
<td>.007</td>
<td>−0.09 (−.05 to −.02)</td>
</tr>
<tr>
<td>Region [British Columbia]</td>
<td>−1.02 (0.63)</td>
<td>−1.64</td>
<td>.10</td>
<td>−0.1 (−.13 to .01)</td>
</tr>
<tr>
<td>Political orientation</td>
<td>−0.39 (0.14)</td>
<td>−2.76</td>
<td>.006</td>
<td>−0.03 (−.06 to −.01)</td>
</tr>
<tr>
<td>COVID-19 threat</td>
<td>0.15 (0.01)</td>
<td>13.53</td>
<td>&lt;.001</td>
<td>0.21 (0.18 to .24)</td>
</tr>
<tr>
<td>Responsibility</td>
<td>7.66 (0.24)</td>
<td>31.90</td>
<td>&lt;.001</td>
<td>.5 (0.47 to .53)</td>
</tr>
<tr>
<td>Trust</td>
<td>0.08 (0.01)</td>
<td>6.04</td>
<td>&lt;.001</td>
<td>.09 (0.06 to .12)</td>
</tr>
</tbody>
</table>

Discussion

At the time of writing, Canada was entering the fourth wave of COVID-19, with case and hospitalization numbers projected to spike in the coming weeks [2,22]. Maximizing vaccination coverage is paramount, but support for public health measures, including physical distancing, masking, staying home while sick, and avoiding crowded indoor spaces, is also critical for limiting the spread of the virus. This is particularly important since some jurisdictions have moved away from mandatory to recommended measures, relying on the public to make adherence decisions. There is an urgent need for effective messaging to increase adherence to public health measures.

Through a randomized online experiment, we tested the effectiveness of five messages featuring different persuasive appeals (deontological vs empathy vs goal proximity vs reciprocity vs identifiable victim) relative to a control message that simply listed the actions participants could take to help prevent the spread of COVID-19. A pretest-posttest design allowed us to assess and compare the change in intentions after exposure to the various messages. The study produced notable insights. First, baseline intentions across all conditions were relatively high (mean 87.18, SD 17.70 on a 100-point scale). Despite our effort to recruit a demographically representative sample, our pool of respondents may have been skewed toward higher compliance. High baseline intentions could also reflect a degree of social desirability bias in the responses. It is worth noting that similarly high levels of self-reported intentions have been observed in prior research [13,21].

Second, exposure to all messages, including the control message, resulted in a small but statistically significant increase in behavioral intentions. Moreover, the message featuring an empathy appeal increased behavioral intentions to a greater extent than the control message. Given how high intentions were to begin with, a small increase should be considered a significant win.

Third, the impact of persuasive appeals on change in intentions depended on how compliant people were in the first place. For those with lower baseline intentions, messages featuring empathy, deontological, reciprocity, and identifiable victim appeals resulted in greater change than the control message. These results are encouraging, as the intended persuasion targets are precisely those who are less compliant with public health measures.

Finally, the study confirmed much of what prior research had found regarding the correlates of public health compliance. Lower baseline intentions were associated with being male, younger, more politically conservative, residing in the Prairies, perceiving lower levels of COVID-19 threat, accepting less responsibility for the well-being of others, and lacking trust in public institutions [49-53]. These results provide a clear and actionable profile of the audiences that need to be targeted to maximize the efficiency of public health campaigns.

While the findings are reasonably informative, it is important to keep the study’s limitations in mind. For instance, the main outcome consisted of self-reported behavioral intentions. Since a gap often exists between intentions and behavior [56], the observed outcomes may not track perfectly with actual behavior. Moreover, as is the case for all studies of this kind, the results...
are likely context-dependent. The same appeals may produce vastly different responses in different countries and at different times, depending on cultural values and the COVID-19 situation on the ground. Thus, it is important not to overgeneralize when interpreting the results.

Importantly, the study used a single brief exposure to the messages, offering a conservative test of the messages’ persuasive power. Future research could investigate whether more frequent exposure or a prolonged exposure period would have a stronger impact. Future research could also test the impact of varying the message format (eg, video vs audio vs print), medium (eg, social media vs traditional media), and source. While the Public Health Agency of Canada is generally a trusted source [53], some groups may respond more positively to other sources (eg, trusted religious and community leaders). Although the focus of this study has been squarely on persuasive appeals, public health campaigns would do well to customize not only the content of the message but also its source, format, and media to maximize its impact across different audiences.

Acknowledgments
We would like to thank the team members at Critical Mass Inc who contributed to participant recruitment. This study was supported by an ImplementAB.digH Program Grant from Alberta Innovates (Grant 202101302).

Authors' Contributions
MM, JLB, RL, MMF, JCB, KC, CC, TT, DAM, and JH conceived and designed the study. MM performed data analysis and wrote the first draft of the manuscript. All authors participated in critical revision of the manuscript and approved the final version. MM is the guarantor of the work and takes responsibility for the integrity of the data.

Conflicts of Interest
DAM reports non-financial support from ISPOR, grants from Canadian Institutes of Health Research (CIHR), Genome Canada, Arthritis Society, and Alberta Innovates; personal fees from Analytica, Illumina, and Novartis. The grants and fees were received during the the timeline of this study but were unrelated to it.

Editorial Notice
This randomized study was only retrospectively registered, explained by authors with the formative nature of the study. The editor granted an exception from ICMJE rules mandating prospective registration of randomized trials because the risk of bias appears low and the study was considered formative. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness, as retrospective registration does not prevent authors from changing their outcome measures retrospectively.

Multimedia Appendix 1
Quota system details (Table S1) and flyers for the persuasion conditions (Figures S1-S5).

Multimedia Appendix 2
CONSORT checklist.

References


42. Cialdini R. The science of persuasion. Sci Am 2001 Feb;284(2):76-81 [FREE Full text]


Abbreviations

- OLS: ordinary least squares
- T1: pretest time point
- T2: posttest time point

©Mehdi Mourali, Jamie L Benham, Raynell Lang, Madison M Fullerton, Jean-Christophe Boucher, Kirsten Cornelson, Robert J Oxoby, Cora Constantinescu, Theresa Tang, Deborah A Marshall, Jia Hu. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 13.02.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Health Care Professionals’ Experiences in Telerehabilitation: Qualitative Content Analysis

Maria Qvarfordt1, MSc; Evalill Nilsson1, PhD; Lina Nilsson1, PhD

eHealth Institute, Department of Medicine and Optometry, Faculty of Health and Life Sciences, Linnaeus University, Kalmar, Sweden

Corresponding Author:
Maria Qvarfordt, MSc
eHealth Institute
Department of Medicine and Optometry, Faculty of Health and Life Sciences
Linnaeus University
Norra Kajplanen 6
Kalmar, 391 82
Sweden
Phone: 46 736120057
Email: maria.qvarfordt@lnu.se

Abstract

Background: The use of digital communication in Swedish health care has increased in an effort to make health care more accessible. At the organizational level, trust in digitalization has stabilized, but a certain degree of skepticism regarding technology appears to exist among health care employees.

Objective: This study aimed to explore health care professionals’ (HCPs) experiences of digital communication with patients and colleagues in a habilitation context.

Methods: Qualitative content analysis was used to analyze data derived from individual interviews.

Results: The results revealed that there were mixed feelings regarding the digital format used at the habilitation center. Although some skepticism remained regarding the digital format, there seemed to be a parallel understanding of the motives and benefits of digitalization. Hence, positive aspects, such as increased health care accessibility, were identified. However, emphasis was placed on the considerations required to make digital consultations appropriate for each patient.

Conclusions: Managing a workday influenced by the balance between digital and physical demands forces HCPs to adjust to the digital format and new ways of working. This requires HCPs to consider whether digital means are appropriate for communication in individual patient-specific cases.

(JMIR Hum Factors 2023;10:e40690) doi:10.2196/40690

KEYWORDS
digitalization; eHealth; habilitation; health care digital encounters; telemedicine; telerehabilitation; HCP; health care professionals; experience; workflow; health care; accessibility; health care employees; perspectives; acceptability

Introduction

Within health care, “habilitation” pertains to helping individuals with congenital developmental disabilities achieve or improve skills and functions necessary for daily living [1,2], whereas “rehabilitation” pertains to helping individuals regain skills, abilities, or knowledge that may have been lost or compromised because of illness or injury or after acquiring a disability [3]. Habilitation services include evaluations, assessments, monitoring, supervision, education, consultation, and coaching. These services can be provided in a physical setting at the habilitation clinic (HC) but also remotely through the use of digital tools (telemedicine) [4]. The latter is part of the ongoing health care digital transformation [5], where digital tools are increasingly used for patient diagnostics, monitoring, and treatment [6], for example, digital consultations [7]. In digital consultations, video can be used as a medium for communicating synchronously [8], but asynchronous communication may occur as well, for example, via text messages [9] and web portals [10]. Although telerehabilitation is an accepted telemedicine subfield [11], telerehabilitation is not commonly used [4]. Rather, it seems to be considered a part of telerehabilitation, although the purpose and requirements may differ. Telerehabilitation has been shown to reduce health care and patient costs [12] and may have the
potential to reduce patients’ time [12,13]. Adding to the benefits of digitally delivered health care, telerehabilitation has also been proposed to increase accessibility [12]. Despite these advantages, challenges remain, such as cases in which digital patient consultations are appropriate or sufficient [14] and addressing prevailing skepticism toward digital tools among health care professionals (HCPs) [12].

There is a lack of evidence on the impact and efficacy of digital tool use in practice, such as for the delivery of higher-quality health care [14,15]. For example, more evidence is needed on how digital communication in health care is best implemented as a replacement for “traditional” patient consultations and on how to run digital consultations most effectively. Is there a limit to when digital consultations would no longer mean an improvement in care but rather a deterioration of care [16]? As an increasingly digitalized working environment offers opportunities to reshape the traditional ways of providing health care [16], possible benefits and disadvantages for patients and HCPs merit further exploration [16], for example, addressing how digital tool use may influence HCPs’ working conditions [14]. Additional questions are raised regarding accessibility, as accessibility in telerehabilitation is associated with how digital tools are designed [17], for instance, with respect to patient inclusivity [18], which is one of the core concerns of habilitation services [3]. Thus, the objective of this study was to explore HCPs’ experiences of digital communication with patients and colleagues in a habilitation context.

Methods

Overview

Qualitative interviews were conducted to explore HCPs’ experiences at an HC in southern Sweden. The interviews were conducted using a data management process modified by Halcomb and Davidson [19]. The following report was inspired by the COREQ (Consolidated Criteria for Reporting Qualitative Studies) guidelines [20].

Study Context

For the purposes of this study, individuals who are in contact with the HC for health care purposes are referred to as “patients.” For digital communication between patients and HCPs, but also between HCPs and other actors, such as other HCPs or other professions at different work sites, schools, or with close contacts—that is, parties involved in a patients’ everyday lives as part of the clinical setting—the term “video consultation” will be used. For digital communication between employees at the HC in organizational settings, such as when attending staff meetings or conferences, the term “videoconferencing” [21] will be used. In both instances, digital communication may be synchronous or asynchronous.

As part of the decentralized Swedish health care system, HCs are organized under the medical or social departments at a regional level [22]. HC employees work in multiprofessional teams that include, for example, nurses, physicians, physiotherapists, psychologists, and speech therapists. Although face-to-face consultations at the clinic were common practice before the COVID-19 pandemic, digital meetings also occurred to some extent, as this particular clinic had implemented video consultations between HCPs and patients before 2020. However, during the study period, all meetings, conferences, and patient visits at this HC were redirected and held via digital formats because of the restrictions imposed by the COVID-19 pandemic.

Sampling Procedure and Participants

The HC in this study comprised several units, 3 of which were eligible for our study. The 2 ineligible units included the unit for children and youth (aged ≤18 years) and the Assistive Technology Center. The inclusion criteria for this study were employment at the HC and contact with patients (aged ≥18 years) and colleagues as part of the participants’ professional workday routine, as described in the Study Context section. The sampling procedure in this study followed a stratified purposeful process [23], as researchers aimed to assemble a heterogeneous group as possible, with respect to participant profession and years active in the profession.

In September 2020, one researcher (MQ) contacted the operations developer of the HC to inform them about the study. The operations developer became our contact person during the study period and contacted the HC management to inform them about the study. When the operations developer received approval from the management that the study could proceed, HC employees were informed about the study via a presentation at a web-based workplace meeting by researchers (LN, EN, and MQ) in October 2020. If interested in participating, employees were requested to correspond with the operations developer, who forwarded potential participants’ contact information to 1 of the researchers (MQ). MQ contacted potential participants via email to provide written information about the study and to book interview appointments. The intention was to use a purposeful stratific strategic sampling. A total of 11 participants signed up for the study, and further sampling was not necessary.

Data Collection

By following an interview guide consisting of semistructured open-ended questions, individual interviews (N=11) were conducted in May and June 2021 by 2 researchers (MQ and LN) via a videoconference. During the interviews, 1 of the researchers (LN) took notes, and after every session, the researchers held a reflexive dialogue about what could be summarized from the interview in line with the analysis method [19].

The interview questions were arranged into 3 themes that were constructed to capture HCPs’ experiences of digital communication with colleagues and patients (Table 1).
Table 1. Interview questions arranged into 3 themes that concerned digital communication in habilitation.

<table>
<thead>
<tr>
<th>Themes</th>
<th>Examples of interview questions</th>
</tr>
</thead>
</table>
| Patient communication and digital patient communication (opening questions) | • Describe in what ways you communicate with patients within your business today regarding patient notifications and follow-ups?  
• What digital tools or facilities do you use to communicate digitally within the habilitation center? |
| Experiences with digital meetings and digital patient encounters | • Keeping digital communication in mind: what opportunities do you believe are offered by the digital format?  
• Keeping possible self-experienced challenges with digital communication in mind: how could the digital encounter with patients be improved? |
| Experiences of digital patient communication: making adjustments because of demands of the digital format | • In what ways do you, as a health care professional, need to adjust your working day because of the use of digital tools such as digital meetings with patients?  
• Describe whether there are any resources saved because of the use of digital communication tools? (eg, time, efficiency, and preparations) |

Before the interviews, participants were informed orally of the study’s purpose in accordance with the informed consent document that was sent via email when scheduling the interview appointment. Only the audio from the interviews was recorded on an external audio-recording device. In addition to providing oral consent at the beginning of the recording, participants also emailed or mailed written informed consent to 1 of the researchers (MQ). Recordings were saved as audio files in a data-secure storage folder, which only the researchers had access to. Two of the interviews were not recorded because of failure of the recording device. However, note-taking from the interview sessions and memos from reflections were documented from these 2 interviews, just as they were for all other interviews.

Data Analysis

The data management process, originally described by Halcomb and Davidson [19], is an iterative reflexive process that integrates the use of verbatim and nonverbatim transcription [19]. We used selective transcription [24], followed by qualitative content analysis, to analyze the data. Finally, a thematic review of the material enabled any necessary changes to be made to codes and themes and aimed for consensus among the researchers (MQ and LN). Codes emerged iteratively during the entire analysis process, and codes were added, deleted, or renamed as the process progressed. During the thematic review of the secondary content analysis, codes and themes were examined to determine the interactions and relationships between them and to ensure that they were a true representation of what had been expressed by respondents.

Ethics Approval and Participation

All interviews were conducted on a voluntary basis, and participants could withdraw from the study at any stage, without being prompted to give an explanation as to why. The interviews were handled under the principle of confidentiality, and the transcripts were pseudonymized. All the participants provided oral and written consent to participate in the study. Professional backgrounds or participants’ specific work sites were not presented in the study to minimize the risk of participants being identified because of the small sample size in relation to the study context. This study was approved by the Swedish Ethical Review Authority (2021-01318).

Results

A total of 11 participants (10 of whom were women) representing the most common professions at the HC participated in the study. Participants had worked for varying periods in their profession, ranging from 1 to 37 years. The interview length ranged between 36 and 56 minutes.

Qualitative Analysis

Two main categories (adapting to the digital working environment and the difference between replacement and complement) and 4 subcategories (cogwheels in the digitalization machinery: importance of support, knowledge, and preparation; workplace-accelerated digitalization: opportunities and challenges for professions; emphasizing the flexibility of digital communication: the digital format as a facilitator; and having case-by-case awareness: digital format suitability) were identified, all emphasizing the overall theme of having to balance a digital-physical workday (Textbox 1).
Adapting to the Digital Working Environment

Overview
An increasingly digital working environment has enabled the HC to continue their everyday work even during ordeals such as the COVID-19 pandemic. However, it was stated that as an HCP, there was a need to adjust to emerging organizational technological frameworks. In addition, support when facing technical issues was thought to be sufficient for some HCPs but insufficient for others. The importance of providing guidance and assistance during the digital workday was emphasized.

Cogwheels in the Digitalization Machinery: Importance of Support, Knowledge, and Preparations
Having to learn and understand digital tools and facilities and constantly being required to have up-to-date knowledge were experienced as challenging by most HCPs. In addition, some HCPs expressed a lack of education on the use of digital tools. Most HCPs stated that manuals and guidelines on how to conduct digital consultations to create a therapeutic alliance with the patient were not available but were desired, as guidelines could ensure that essential elements in the communication process would not become lost, thus helping to ensure the quality of communication.

Technical support and advice were thought to be well functioning, as almost all HCPs were satisfied with phone support and stepwise guides for digital tool use. Some HCPs experienced support from “super users” (colleagues with cutting-edge expertise) as very useful. However, hands-on guidance regarding technical issues was expressed as inadequate by few HCPs.

Regarding the adequacy of digital communication, most HCPs expressed the need to establish certain prerequisites for digital use within the organization; employees’ digital competence and capabilities (pertaining to digital methods and tools) need to be sufficient to function optimally. Adequate patient digital literacy, that is, competence from the initial log-in step; the ability to manage digital tools well; and digital capabilities such as having a stable internet connection with adequate video and audio quality and up-to-date software and digital devices were understood by most HCPs to be vital to their work.

Workplace-Accelerated Digitalization: Opportunities and Challenges for the Professional Role
Most HCPs expressed that when meetings with colleagues were held in a digital-physical setting (blending digital participation and physical participation), it led to less interaction between attendees, occasionally leaving some attendees feeling that they had not been part of the meeting. When attending digital conferences, it was expressed that it was difficult to visually raise hands and share thoughts, partly because of a fear of “social clumsiness,” which resulted in feelings of frustration. Conversational turn-taking was experienced as time consuming, yet it was considered necessary in some groups.

It was perceived that digitalization has developed into a natural part of how health care could (and maybe should) be delivered. Some HCPs experienced gratification in having been quickly “thrown into” an increasingly digitally managed organization because of the COVID-19 pandemic, as this enabled them to assess how digital workdays might be arranged in the future. However, some HCPs observed a certain degree of skepticism toward digitalization as part of quality management among colleagues, but the acceptance of digitalization seemed to be dominant.

Most HCPs experienced increased cognitive fatigue and tiredness because of the increasing number of digitally held conferences and consultations. The need to learn new technologies and working routines was constantly ongoing, which could also contribute to cognitive fatigue. Technical errors were a source of frustration and stress, and some HCPs mentioned that technical errors caused a sense of hopelessness.

Some HCPs experienced working from home as positive, as this type of setting minimized stress and eased everyday life coordination, for instance owing to time savings. The increasing number of digital consultations at the HC allowed a considerable portion of planned activities to be maintained, thus enforcing accessibility despite the COVID-19 pandemic, and most HCPs stated that the digital working day was desirable as a way to continue working in the future. However, the digital working day was considered a suitable complement to more traditional ways of working that relied heavily on physical presence, as the more traditional way of approaching health care was stated as important because of the varying levels of digital literacy among patients and providers.

The Difference Between Replacement and Complement

Overview
The HCPs found digital consultation to be a suitable complement in many cases and settings. However, consideration of the suitability of the digital format to the specifics of each particular case was expressed as crucial. The importance of having
additional, more traditional ways of contacting HCPs was emphasized.

**Emphasizing the Flexibility of Digital Communication:**

**The Digital Format as a Facilitator**

Most HCPs pointed out that individuals diagnosed with neuropsychiatric disorders might benefit from communicating digitally with HCPs; as attending digital consultations from their home environment might be comforting, enabling a dedramatizing, informal setting. Digital consultations were described as most suitable for individuals with anxiety or other social, psychiatric, or mental health impairments owing to the flexibility and accessibility features of the digital format. Digital consultations were considered to not be as mentally exhausting and intense for the patient as face-to-face consultations.

It was expressed that a digital consultation for an initial patient contact could promote future face-to-face consultations at the clinic, as digital communication was thought to encourage the establishment of the HCP-patient alliance in some cases. In addition, some patients were considered to be more open when they were met through a digital format, which was mainly acknowledged when patients were located in a home environment while they were attending the consultations. Attending from familiar environments was perceived as giving patients a sense of safety. This was thought to be especially applicable to individuals diagnosed with neuropsychiatric disorders. Some HCPs experienced digital consultations as reducing both the rescheduling and cancelation of appointments.

When arranging digital conferences, some HCPs found it easier for different parties to agree on a consensus suitable time for digitally held meetings than for meetings held in a physical setting. A few HCPs mentioned that the nervousness that they sometimes experienced before meetings with colleagues was reduced, as digital conferences reduced the feeling of being “reviewed” by other meeting attendees. It was considered easy and thus possible to be able to connect to various meetings digitally (with parties such as patients and stakeholders), also acknowledging the environmental and time-saving benefits associated with decreased traveling. The digital format for meetings was considered to enable a more efficiently planned workday.

**Having a Case-by-Case Awareness: Digital Format Suitability**

Almost all HCPs expressed concerns about individuals with insufficient digital literacy, as this was thought to be related to health care accessibility. Some concerns were raised about individuals with poorer health, as a few HCPs reported that poor health might impair cognitive ability, which in turn could negatively influence the patient’s capability of using digital tools. The importance of additional means of contacting HCPs in addition to formats was thus stressed, as health care accessibility was considered crucial in affecting individuals with insufficient digital literacy.

Digital consultations occasionally caused some HCPs to question and worry about their ability to make accurate assessments digitally, as it was difficult to acknowledge “vital signs” of importance to the clinical assessment, such as visual indications of abuse, or ill-health, such as pale skin and skin rashes. Ensuring patient safety was considered complicated because of the impaired ability to make a holistic evaluation of individuals based on bodily expressions that may indicate restlessness and nervousness during digital consultations with patients. In addition, most HCPs stated that the digital format impaired interpersonal interaction, impacting nonverbal cues such as bodily expressions and gestures, as these were diminished to some extent. Important elements such as being able to better observe or better understand individuals’ interaction and communication with their surroundings were considerably affected and were considered by some HCPs to have a potentially negative effect on the HCP-patient alliance.

A few HCPs expressed concerns about the suitability of prescribing medication based on digitally conducted assessments and also raised concerns about being unable to conduct medication follow-ups, as the digital format made it difficult to make before and after comparisons of patient restlessness or other factors that may be affected by medication, because of the diminished availability of nonverbal cues.

Some HCPs experienced that some forms of consultations required a physical meeting space (such as the HC) to create a safe and comfortable environment for patient communication. Nearly all HCPs emphasized the need to assess whether a digital consultation was appropriate for each case, stating that the digital format is not suitable for everyone or suited to everyone or for every purpose. For example, digital consultations were considered to be unsuitable for patients who easily become unfocused or exhausted during the meeting. According to all HCPs, the inability to illustrate while interacting with patients and the use of whiteboards as examples of communication-promoting tools were experienced as negatively affecting communication quality. Almost all HCPs felt that digital consultations did not enable detailed communication or promote in-depth dialogue to the same extent as in a physically held consultation. Some HCPs expressed that this was affected by the reduced dynamic between meeting attendees and the fact that, although preferred by some, others need physically held consultations to be able to open up in more when communicating.

A few HCPs experienced that digitalization should not only be considered for individuals who are already skilled; even digital novices might benefit from the digital format if they learned to take advantage of it, in accordance with sufficient digital literacy. Putting the patient’s perspective first in managing their care was considered central to this issue, as some HCPs also stated that health care employees need to be encouraging and offer help and support in this transformation toward an increasingly digitalized context. Furthermore, it was the view of HCPs that patients often needed assistance in their initial forays into digital tools and facilities, and supporting individuals in this way was sometimes considered time consuming for some HCPs, but still worth the effort, as promoting more digital skills in both patients and providers was perceived as enabling more efficient management of health care.

https://humanfactors.jmir.org/2023/1/e40690
Discussion

Principal Findings

The main results of the study showed that maintaining activities within the HC owing to organizational development (prompted by the COVID-19 pandemic) required HCPs to be part of an increasingly digitally managed workplace. HCPs were constantly influenced by the overall theme of this study: having to balance between the digital and physical work activities during a workday and having to constantly review digital communication as a complement versus as a replacement. Thus, the digital format was not perceived as suitable for everyone, at every time, or for every purpose, although digital communication was often perceived as a useful complement in many cases and settings among the HCPs in this study.

The results of this study will be discussed from 2 viewpoints: the patient contact perspective and the workplace perspective, with an additional focus on collegial interactions.

The Patient Contact Perspective

In this study, nearly all HCPs emphasized the importance of acknowledging the difference between replacement and complement regarding the use of digital consultations and of acknowledging its varying suitability. For example, individuals diagnosed with neuropsychiatric disorders were considered to benefit from the digital format when communicating with HCPs, as individuals diagnosed with neuropsychiatric disorders might feel more comfortable attending digital consultations from an already well-known environment as opposed to attending face-to-face consultations at the clinic, which might be less familiar. Digital consultations could also enable a dedramatizing, less stressful, and more comfortable meeting environment for patients compared with face-to-face consultations. In addition, previous research suggests that patients experience a chat function for health care communication purposes as more considerate of their conditions compared with face-to-face communication [25]. It has also been previously suggested that interaction with HCPs via a digital platform (including chat-based features and phone calls) might encourage patients to open up more easily and share their thoughts more owing to increased engagement from HCPs [26].

In contrast, some HCPs in this study found it difficult to foster in-depth communication during digital consultations. Previous research suggests that health care delivered digitally will affect the HCP-patient alliance in different ways [27], as the use of nonverbal cues such as bodily expressions and gestures stimulates the establishment of the HCP-patient alliance [28]. Nonverbal cues offer insight into the underlying unstated concerns and emotions and also support the reinforcement or contradiction of verbal communication [29]. In addition, most HCPs experienced that interpersonal interaction was diminished during digital consultations, as nonverbal cues were minimized. The negative effect of the digital format on these aspects of communication has been previously suggested [30], which implies that over time, these aspects could negatively affect treatment [31,32]. In contrast, it was previously suggested that telemedicine provides similar [33] or better outcomes than conventional face-to-face health care [34], but always assessing the suitability of a digital consultation in each individual case is thought to be of utmost importance [35]. The latter supports the findings of this study, as HCPs stressed the importance of always carefully considering the appropriateness of digital communication from the perspective of each patient.

Almost all HCPs experienced that there were no easily accessible guidelines provided by their organization about how to best conduct digital meetings. Previous research emphasizes the use of guidelines and understanding a patient’s previous digital experiences and digital literacy [36]. Similarly, the results of this study indicate that digital literacy and sufficient digital capabilities are required when conducting a digital consultation. Along with digital capability requirements, such as managing digital methods and tools [37], HCPs experienced a need to adopt the role of patient educators in digital matters. Previous research implies that it may be challenging for individual clinics to handle digital transformational processes, leaving HCPs to improvise and individually evaluate whether health care can be safely delivered [38]. In addition, previous research also suggests that, to prevent digital exclusion, HCPs should offer information, encouragement, or tools for patients [39], which is in accordance with the findings suggesting that tailored education for individuals is needed within the telehealth domain [40].

Most HCPs highlighted the lack of opportunities to complement verbal communication, such as drawings to assist verbal communication and the use of whiteboards. These communication-promoting tools were normally used at the HC and were considered to highlight and concretize verbal HCP-patient communication. Technologies are constantly undergoing development, modern videoconference tools provide virtual white boards as integrated parts of the software [41], and previous research suggests that virtual whiteboards support children’s collaborative communication abilities in a classroom setting with reference to the communication process and use of a “learning resource” as part of integrated system of spoken dialogue and nonverbal communication [42]. Hence, the use of digital whiteboards as part of an integrated videoconference tool might constitute a possible alternative to more traditional communication-promoting tools such as paper and pencil or physical whiteboards.

To tackle these challenges, guidance [43], support for developing and maintaining HCP-patient alliances [44], and methods of how to conduct best practices in the context of increasing digitalization need to be addressed [45]. Raising awareness of the risks of digitalization of health care is vital, as is adopting techniques for sustainable digital clinical relationships [44]. In addition, previous findings indicate that HCPs are generally positive about implementing therapeutic initiatives digitally. Although attitudes might have been influenced by previous experiences related to the clinic and changes within and previous digital format use, feelings of fatigue, incompetence, and insecurity were expressed, as were experiences of having reduced physical contact with the patient [46].
The Workplace Collegial Perspective

Some HCPs experienced frustration with technological failures and having to rely on their own insufficient expertise, which conforms to previous findings that digitalization is not always implemented in ways that take digital competence into account. This is an important point to take seriously, as having insufficient technological competence might cause frustration among HCPs when trying to adopt new technologies [47]. In addition, the accelerated changes forced by the rapid introduction of digitalization were not experienced as being followed by sufficient professional training and guidance. This has been previously demonstrated within the telemedicine field [36] and might be important, as HCP competence and willingness are crucial factors for successful health care digitalization [48]. Furthermore, some of the recent rapid uptakes of digital tools in Swedish health care have been accepted, as this aid was prompted by exceptional circumstances. Considering these digital tools as possible benefits in health care seems to depend on a large extent on individual and organizational aspects, rather than on technological aspects alone [49]. However, HCP education regarding how to conduct digital consultations has not been extensively realized [14], although it has been previously suggested that HCPs need training regarding newly implemented tools [50], for example, in learning how to efficiently communicate digitally in clinical matters [47,51].

To harmonize digitalization within the health care sector, assessing and improving HCPs’ digital competencies might be an initial step in incorporating digitalization effectively into clinical practice. This is important from an organizational development perspective, as providing the best possible health care for patients requires HCPs to develop profound knowledge and skills in relation to new working ways prompted by digitalization [52]. Previous research also pinpoints that employee’s social relations might influence behavior and acceptance of novel technologies [53], which agrees with the findings of this study, seen in the positive experiences that some HCPs experienced regarding receiving help from colleagues with special competence, the so-called “superusers.”

Although situations such as the COVID-19 pandemic might stimulate the rather prompt and less prepared implementation of telehealth solutions within health care [54], the HCPs in this study experienced being hastily thrown into the digitally working climate as quite positive, having the advantage of being able to pilot future possible working methods. Although the rapidly emerging situation seemed to be slightly amorphous at first, the HCPs emphasized the flexibility of the digital format. This corresponds to previous findings indicating that quickly adopting digital solutions may catalyze telehealth development within organizations [38].

As part of an increasingly digital working environment, working from home was perceived as advantageous owing to the increased flexibility experienced by most HCPs. It was considered easy and thus positive to be able to connect to various meetings digitally, which was also expressed as positively related to the environmental benefit attributed to decreased business travel, which has been previously reported [55,56].

Most HCPs were relatively positive toward digitalization as a future feature of the workplace, embracing its presence and accepting digital solutions as potential answers to the ongoing and progressive health care management transformation. In the literature on digital tool assistance, the term “useworthy” emerged, aiming to demonstrate not only the usefulness of a technology but also to show its value as it meets the high priority needs of the users [57].

Digital conferences for workplace events such as staff meetings were experienced as convenient by most HCPs because of resource savings, scheduling advantages, and increased accessibility. In accordance with these findings, previous research suggests that telehealth technology use enables HCPs to more easily share information and collaborate in patients’ treatment [58], improving interdisciplinary collaboration [59], thus helping to overcome collaboration barriers [60] and ensure continuity of care.

Always being up-to-date regarding software and holding nearly all conferences and consultations digitally were perceived as tiring to some HCPs. Furthermore, our findings imply that this increased tiredness prevails in part because of an increasingly digitalized working environment, which has previously been explored in terms of workplace digital fatigue [61,62], more specifically mentioned as “videoconference fatigue” because of increased participation in digitally held meetings [63]. However, it is important to emphasize that for any meeting to be successful, in-person and digital meetings require adequate preparations, such as sharing relevant documents and agendas before the meeting [41]. Although videoconference fatigue conceptually belongs to the more common construct of “work fatigue,” the 2 concepts differ, as work fatigue is mainly associated with workplace demands in general, such as work overload and time constraints [64]. These features may also apply to videoconference fatigue, but as a concept, videoconference fatigue is more specific than the general causes of work fatigue and could conversely be generated as a consequence of single events. For example, being active digitally imposes avoidance of technology-based distractions, while also calling for greater attention to be paid owing to fewer available nonverbal communication cues [65]. This is in accordance with the findings of this study, implying that most HCPs experienced that cognitive fatigue, besides tiredness, occurred because of an increased number of digitally held conferences and consultations.

In this study, closely related to having to balance the digital-physical workday, digital solutions used within the organization correspond to the value of technology and thus its “worth,” as using technology fulfills the clinic’s needs, maintaining valuable HCP-patient contact while keeping the focus on patients. As a knowledge gap seems to exist regarding habilitation services overall, our purpose was to further explore the field of habilitation explicitly. We would also like to highlight that more research is needed within the domain of telehabilitation, both from an HCP and a patient perspective, with respect to the multifaceted profile of HC clients. Moreover, to apply a broader perspective, a macrolevel insight would be useful to obtain, for example, by inquiring into the managerial viewpoint of digitalization in (clinic-specific) health care further.
exploring the use of digital tools and digitization in health care in general from a managerial perspective, implementation strategies may be worth exploring. This may be particularly relevant considering the importance of sustainable health care management and development. Emphasizing core values in health care, that is, health care should undergo constant quality improvements [66], with a focus on aspects such as patient engagement, patient-centeredness, and health literacy [67], future research may focus on digital tool use in matters of clinical assessments, more explicitly regarding how best to conduct clinical assessments sufficiently via a digital format as a part of digital consultations. On the basis of the results of this study, we propose that future research may target whether digital formats mediate a holistic view of an individual and adequately provide HCPs with sufficient amounts of clinical information to take further actions, for instance, regarding drug prescriptions or further referrals in health care. This question is of great significance to patient safety, as digital health care consultations might be less appropriate for some patient conditions or when technological shortcomings suddenly interrupt consultations [68]. The issue of patient safety is also very important for HCPs, whose professional role includes showing empathy and compassion as well as professional integrity [69].

Methodological Considerations

In this study, the HC’s operations developer was contacted and asked to arrange initial contact with potential participants. Using the operations developer at the HC for participant recruitment could be associated with an ethical risk [70], for instance, as some participants may have felt prompted to participate, thereby not fully conforming to the rule of voluntary participation. There is an additional risk of biased sampling, as the person assisting in the recruitment process may easily become too helpful, wanting to recruit participants who they think can provide suitable answers [70]. However, participants were not recruited directly by the operations developer; they were asked to contact 1 of the researchers via the operations developer if they were interested in participating after being given additional information about the study at the web-based workplace meeting. Furthermore, self-selection bias may also occur because of personal engagement in the digitalization process at the HC, in accordance with participants’ specific characteristics; thus, some HCPs might have been more likely to take part in the study than others. However, self-selection bias is difficult to avoid completely in interview-based research because voluntary participation is central to ethical good practices [71]. Furthermore, people who volunteer in this way are more likely to have things on their mind, positive aspects as well as negative, that they would like to convey; hence, they are probably better informants in an interview study.

All data were collected using videoconferencing. Previous research supports the use of digital methods for qualitative data collection because of their cost-effectiveness, security options, ease of use [72], and relaxed environment, which can occasionally foster deeper conversations [73]. In contrast, feelings of videoconference fatigue may have potentially affected participants’ willingness to participate in a digital interview [63]. Moreover, interviews conducted in a digital format might exclude some populations owing to varying levels of digital literacy [72], and the digital format also limits researchers’ opportunities to fully observe the full range of body language and nonverbal communication [74]. Selective transcription was used in the analysis process, as suggested by Halcomb and Davidson [19], following their guidance on the analysis process. Using audio recordings along with memo writing might help assure methodological accuracy in terms of credibility [75], as memo writing was complemented by listening to recordings when conducting the 2-step content analysis [19]. The use of written memos conducted either during an interview or immediately afterward has been suggested to be superior to only the use of audio recordings transcribed verbatim in terms of enhanced trustworthiness [76]. The fact that the 2 researchers who conducted the interviews also reached a consensus during the analytical phase further contributed to this study’s trustworthiness [77]. It can be concluded that data saturation [78] was reached when looking through and reflectively discussing the notes shortly after the last 3 interviews. At this point, the researchers estimated that no further information would be obtained [79]. To have most of the study population defined by only 1 gender might, however, affect the point at which data saturation was reached. However, it has previously been shown that studies with relatively homogenous study populations may reach reliable saturation at quantities similar to those in this study [80].

Conclusions

Managing a workday influenced by the balance between digital and physical demands forces HCPs to adjust to the digital format as part of an increasingly digitally managed workday. Being aware of digitalization as a workplace development process and constantly having to adapt to changing demands (considering that digital formats are not suitable for every patient encounter) is a complex yet required task. Driven by professional values such as putting patient care first, negotiating the pros and cons of health care digitalization is a constantly evolving and challenging process.

Therefore, having to balance work that bridges both the physical and digital work activities in times of rapid organizational development fueled by digitalization spurs the need to acquire knowledge on the adoption of new ways of working. In an HC setting, the introduction of digital tools to increase knowledge and the possibility of tailoring patient visits to different patient populations, such as individuals diagnosed with neuropsychiatric disorders, is likely to complement more traditional ways of practicing medicine.
Acknowledgments
The authors wish to express their gratitude to the participants of this study for sharing their experiences. The authors would also like to thank the members of the project group, particularly Maria Borg, for their support. This study received financial support from the Kamprad Family Foundation. The funders were not involved in the study.

Conflicts of Interest
None declared.

References

https://humanfactors.jmir.org/2023/1/e40690


Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Studies
HC: habilitation clinic
HCP: health care professional

©Maria Qvarfordt, Evalill Nilsson, Lina Nilsson. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 19.04.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Involving Older People With Frailty or Impairment in the Design Process of Digital Health Technologies to Enable Aging in Place: Scoping Review

Emilie Kauffeldt Wegener1, MA, MSc; Jenny M Bergschöld2, PhD; Carly Whitmore3, PhD; Marjolein Winters4, MSc; Lars Kayser1, MD, PhD

1Department of Public Health, University of Copenhagen, Copenhagen, Denmark
2Department of Health, SINTEF Digital, Oslo, Norway
3School of Nursing, McMaster University, Hamilton, ON, Canada
4Smart Homes, Eindhoven, Netherlands

Corresponding Author:
Emilie Kauffeldt Wegener, MA, MSc
Department of Public Health
University of Copenhagen
Øster Farimagsgade 5
Copenhagen, 1353
Denmark
Phone: 45 26202168
Email: emilie.wegener@sund.ku.dk

Abstract

Background: With an increase in life expectancy globally, the focus on digital health technologies that can enhance physical and mental health among older people with frailty and impairment has increased. Similarly, research interest in how digital health technology can promote well-being and self-management of health in older age has increased, including an increased focus on methods for designing digital health technologies that meet the various medical, psychological, and social needs of older population. Despite the increased focus, there remains a necessity to further understand the needs of this population group to ensure uptake and to avoid introduction of additional challenges when introducing technologies, for example, because of poor technological design. The scope is limited to digital health technologies meant to enable older people with frailty and impairment to age in place.

Objective: In this study, we aimed to explore how older people with frailty and impairment are involved in various parts of the design processes of digital health technologies and identify gaps or neglected steps in a user-involving design process. This included a focus on recruitment strategies, contributions, and methods used to address the perspectives, needs, and desires of older people with frailty and impairment in the development of digital health technologies.

Methods: A scoping review was conducted in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) reporting from February 2021 to April 2021. Literature searches were conducted in PubMed, Scopus, Embase, and IEEE using a search string covering the concepts of health technology, older people, frailty and impairment, user-centered design, and self-management.

Results: In total, 1891 studies were imported for screening from the initial search. A total of 22 studies were included in this review after full-text screening and manual search. Invitation through partners was the most reported recruitment strategy to involve older people with frailty and impairment in the design process of digital health technologies. Furthermore, they were commonly involved in the final evaluation of the development process. Three main gaps identified were the use of outreach approaches to recruit older people with frailty and impairment in the design process of digital health technologies, description of the value of involvement and outcome of the contribution of participants, and knowledge regarding involvement in all parts of the design process.

Conclusions: Although there is literature on methods for involving older people with frailty and impairment in the design of digital health technology, there is little methodological dialogue on the nuances of how different methods for involvement relate to and shape the outcome of the development process.
The conduct of medical device clinical studies, including obtaining recommendations on patient engagement in the design and development of products and services with the involvement of end users [19]. These guidelines and recommendations for involving end users in the development of health technology support this work and also demonstrate the increased global attention on this important work.

Although it has been widely accepted to involve end users in the design of new technologies and an increased focus on user involvement among certifying bodies is emerging [20,21], there are no standardized requirements or guidelines on how to involve end users [22]. Also, the academic debate on appropriate methods that involve older people who are physically or cognitively impaired or otherwise understood to be frail and allow them to express their needs and desires vis-à-vis technology is lagging. Consequently, there is a need for further knowledge about how to involve older people with frailty and impairment in the design of technology to ensure that their needs and desires are addressed and to better understand what they find meaningful to increase the likelihood of technology adoption.

**Digital Health Technologies**

In this scoping review, the focus is limited to how older people with frailty and impairment are involved in the design process of digital health technologies. Digital health technologies are defined by the Food and Drug Administration as “the use of computing platforms, connectivity, software, and sensors for health care and related uses” [7]. In this review, we include eHealth and its underlying terms in our understanding of digital health technologies. “eHealth” is defined by WHO as an umbrella term covering the general use of technologies for health care–related processes, including mobile health, the use of different mobile-based solutions, telemedicine, remote clinical services, and telehealth covering both remote and nonremote clinical services [23].

**Aim**

The identification and application of purposeful methods for involving older people with frailty and impairment in the innovation and implementation of digital health technology may be a promising means of ensuring that the technology can fulfill its purpose of enabling such older people to age in place with dignity and on their own terms.

Against this background, this review aimed to explore how older people with frailty and impairment are involved in various aspects of design processes of digital health technologies. This was done to identify gaps or neglected steps in a user-involving design process. This included a focus on recruitment strategies, contributions, and methods used to address the perspectives, as well as the appropriateness of technologies and the end users' experience with them.
well as the needs and desires of older people with frailty and impairment in the development of digital health technologies.

To this end we pursued the following research questions:

1. What are the characteristics of the participants included in the design processes and how are they recruited?
2. Based on the objective of this study, what are the outcomes? What was the technology developed and how did the participants contribute?
3. What kinds of methods and activities have been used to involve older people with frailty and impairment and when were they involved during the development process?

Methods

Overview

The scoping review was conducted as part of an EU-funded collaborative project between the EU and Canada called Smart Inclusive Living Environments (SMILE). The SMILE project is working to support aging in place using eHealth solutions with the aim of enabling older people to live an independent and active life, irrespective of frailty and physical or cognitive impairments, using new technologies developed with and for them. Throughout this review, references will be made to “older people” and “health technologies,” for this work, which refers to older people with frailty and impairment and new digital health technologies, respectively.

We conducted a scoping review following the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) to synthesize knowledge; map existing evidence; and identify concepts, theories, sources, and knowledge gaps [24].

Eligibility Criteria

Inclusion Criteria

Studies were eligible for inclusion if they were published in English and in peer-reviewed journals, including conference papers, in any year as identified by our search strings. Eligible studies had to involve >65-year-old people with frailty and impairment. This includes individuals with cognitive decline or deterioration (eg, dementia), cognitive dysfunction (eg, aphasia), neurocognitive impairment, motor dysfunction (eg, stroke and ataxia), and physical and mental frailty or vulnerability. This information could be self-reported in the study. Furthermore, studies were eligible if perspectives, needs, and desires of older people with frailty and impairment were expressed and included in the development process of a digital health technology. Studies had to include a description of the development or design of digital health technologies. This aligned with the need to understand the methods of involvement in the design of digital health technology.

Exclusion Criteria

Table 1 provides an overview of the exclusion criteria. Review articles were excluded to avoid redundancy with respect to the original articles included in the review. Case reports, abstracts, and conference proceedings presenting preliminary data were excluded. Thus, full-text articles published with respect to conferences were not excluded.

The scope is limited to the use of digital health technologies meant to enable older people to age in place. Therefore, studies addressing the development of everyday technologies (eg, electrical appliances, technologies for indoor climate regulation, vacuum cleaners, jar openers, and electric curtains) in general products that are not used for specific health issues were excluded.

The following studies were also excluded: effect studies, such as those that only evaluated user experience or implementation (ie, acceptance, feasibility, effectiveness, and efficacy); and studies in which the methods of involvement were not reported or the involvement was not in the context of development of digital health technologies for aging in place.
In Covidence, 1.12% (30/2675) more duplicates were removed. Review software tool developed by the Cochrane Community. Mendeley before importing to the Covidence database [25], a total of 29.35% (784/2675) duplicates were removed using.

A total of 2675 studies were obtained using the search strings for PubMed, Scopus, Embase, and IEEE. 922 (34.47%) studies were obtained using the search string for PubMed; 1753 (65.53%) studies were

Obtained from Scopus and 0 (0) from Embase and IEEE. In total, 3 medical subject heading terms were included in the search string for PubMed: Telemedicine established in 1993, Cognitive Dysfunction established in 2012, and aged established in 1966.

Filters for the search strings included full-text availability and English language and excluded case reports, conference proceedings, abstracts, and nonpeer reviewed articles.

Studies in which methods cannot be clearly identified were included in the search string for PubMed: Telemedicine established in 1993, Cognitive Dysfunction established in 2012, and aged established in 1966.

Information Sources and Search

Literature searches were performed in PubMed, Scopus, Embase, and IEEE. After the initial screening process, additional identification of relevant studies was performed using following two strategies: (1) identification of relevant studies in the reference lists of the screened studies and (2) input from experts in a workshop, thereby an additional 11 studies were identified; of these, 1 study met the inclusion criteria.

In PubMed, medical subject heading terms were included (search string for PubMed is listed in Multimedia Appendix 1), whereas in other databases (Scopus, Embase, and IEEE) keyword search was conducted (search strings for Scopus, Embase, and IEEE are listed in Multimedia Appendix 1). In total, 3 medical subject heading terms were included in the search string for PubMed: Telemedicine established in 1993, Cognitive Dysfunction established in 2012, and aged established in 1966.

Filters for the search strings included full-text availability and English language and excluded case reports, conference proceedings, abstracts, and nonpeer reviewed articles.

Screening

In total, 2675 studies were obtained using the search strings for PubMed, Scopus, Embase, and IEEE; 922 (34.47%) studies were obtained using the search string for PubMed; 1753 (65.53%) was obtained from Scopus and 0 (0) from Embase and IEEE. A total of 29.35% (784/2675) duplicates were removed using Mendeley before importing to the Covidence database [25], a review software tool developed by the Cochrane Community.

In Covidence, 1.12% (30/2675) more duplicates were removed based on the title, year, volume, and author. Duplicates were verified by the authors and removed, leaving 69.57% (1861/2675) studies for screening.

The studies were reviewed using Covidence [25]. After title and abstract screening, 63.1% (1688/2675) studies were excluded, with 10.25% (173/1688) studies then assessed for full-text eligibility. After full-text screening, 2.19% (37/1688) studies were included (Figure 1). During the data extraction phase of the review, 0.95% (16/1688) studies were further excluded because they did not meet the inclusion criteria (eg, only reporting on household appliances or design of future homes or lack of reporting on the involvement of older people in the design process). In addition, 1 study was identified by SMILE project partners in a workshop and was included in the review.

In total, 12.7% (22/173) articles were included. The screening process is illustrated in a PRISMA-ScR flow diagram (Figure 1). The first 4 authors conducted the initial screening by titles and abstracts, and the full-text screening was conducted on a first-to-come basis. All screenings (title, abstract, and full text) involved 4 authors, and each article screening included 2 reviewers. When in doubt about the eligibility of an article, all 5 authors discussed the evaluation. The full-text articles and those retrieved from the manual search and workshop were extracted by the authors EKW, LK, CW, and JMB. All authors participated in the synthesis and presentation of the findings.

Of the 22 studies, 9 (41%) studies had included some population aged >65 years. Because most of the population in these studies was >65 years of age, all the authors decided to include those

### Table 1. Exclusion criteria.

<table>
<thead>
<tr>
<th>Exclusion</th>
<th>Label</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Papers that have not been peer-reviewed</td>
<td>Not peer-reviewed article</td>
<td>Also include case reports, conference proceedings, and abstracts.</td>
</tr>
<tr>
<td>Full text not available</td>
<td>Full text not available</td>
<td>N/A a</td>
</tr>
<tr>
<td>Exclude theoretical papers, that is, no involvement of population (eg, research protocols and theoretical papers)</td>
<td>Protocols</td>
<td>Research (protocols) that is planned but not executed.</td>
</tr>
<tr>
<td>If data cannot be clearly identified for the age group or subgroup of &gt;65 years</td>
<td>Population not &gt;65</td>
<td>N/A</td>
</tr>
<tr>
<td>If data cannot be clearly identified for the group or subgroup of older people with frailty or impairment</td>
<td>Population not frail or impaired</td>
<td>Definition of frail and impaired; individuals with following disorders: cognitive decline (eg, dementia), cognitive dysfunction (eg, aphasia), neurocognitive impairment, motor dysfunction (eg, stroke, ataxia), frailty, vulnerability (not only social vulnerability).</td>
</tr>
<tr>
<td>Exclude studies that address the development of everyday technologies such as electrical appliances and technologies for indoor climate regulation and vacuum cleaners. In general, products not used for specific health issues</td>
<td>Everyday technology</td>
<td>This group should be revisited after first round.</td>
</tr>
<tr>
<td>Studies in which methods cannot be clearly identified</td>
<td>Does not report methods of involvement</td>
<td>Definition of involvement: end user’s perspectives, needs, and desires are expressed and included in the development process to an extend beyond focus groups and classical participatory design.</td>
</tr>
<tr>
<td>Involvement not for the purposes of development</td>
<td>Involvement not for the purposes of development</td>
<td>Definition of development: development or design of new innovative technologies.</td>
</tr>
<tr>
<td>Exclude effect studies</td>
<td>Effect studies</td>
<td>Articles that only evaluate user experience or implementation, that is, acceptance, feasibility, effectiveness, efficacy, and so on.</td>
</tr>
</tbody>
</table>

aN/A: not applicable.

### Information Sources and Search

Literature searches were performed in PubMed, Scopus, Embase, and IEEE. After the initial screening process, additional identification of relevant studies was performed using following two strategies: (1) identification of relevant studies in the reference lists of the screened studies and (2) input from experts in a workshop, thereby an additional 11 studies were identified; of these, 1 study met the inclusion criteria.

In PubMed, medical subject heading terms were included (search string for PubMed is listed in Multimedia Appendix 1), whereas in other databases (Scopus, Embase, and IEEE) keyword search was conducted (search strings for Scopus, Embase, and IEEE are listed in Multimedia Appendix 1). In total, 3 medical subject heading terms were included in the search string for PubMed: Telemedicine established in 1993, Cognitive Dysfunction established in 2012, and aged established in 1966.

Filters for the search strings included full-text availability and English language and excluded case reports, conference proceedings, abstracts, and nonpeer reviewed articles.
studies and thus contributes important information about the design of new technologies for the age group of people >65 years.

Results

Overview

The following sections present the results of the data extraction. In the first section, a description of the general characteristics of the studies is provided. The second section presents the results related to the characterization of older people with frailty and impairment involved in the design processes and applied recruitment methods. The third section addresses the outcome of the study, the type of digital health technology developed, and contribution of older people with frailty and impairment to the design process. The fourth section addresses the methods and activities used to involve the older people, as well as time point of involvement during the development process.

General Characteristics of the Included Studies

The 22 studies included in this scoping review were published between 2009 and 2020, indicating that the practice of involving and focusing on how to involve older people with frailty and impairment in the design and development of digital health technologies is an increasing field. The included studies were not limited to a single geographic location. Geographic locations included the WHO Region of the Americas (5 studies from the United States, 1 from Chile, and 1 from Canada), the WHO European Region (2 studies from the United Kingdom; 3 from the Netherlands; and studies from Portugal, Germany, Italy, Finland, and Sweden), the WHO Western Pacific Region (1 study from Malaysia), and the WHO African Region (1 study from South Africa).

Table 2 provides an overview of the population groups included in these studies. Descriptions of the populations show a large representation of somatic conditions, cognitive conditions (eg, dementia and risk of cognitive decline), or a combination of both.
Table 2. Overview of conditions represented in studies (N=22).

<table>
<thead>
<tr>
<th>Author and year</th>
<th>Study population condition</th>
<th>Value, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gövercin et al [26], 2010</td>
<td>Risk of falling</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Hakobyan et al [27], 2015</td>
<td>AMD\textsuperscript{a}</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Willard et al [28], 2018</td>
<td>Risk of cognitive decline</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Bogza et al [29], 2020</td>
<td>Mild cognitive impairment</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Hassan et al [30], 2017</td>
<td>Dementia</td>
<td>2 (9)</td>
</tr>
<tr>
<td>Kerkhof et al [31], 2019</td>
<td>Parkinson disease</td>
<td>2 (9)</td>
</tr>
<tr>
<td>de Barros et al [32], 2013</td>
<td>Patients with heart failure</td>
<td>3 (14)</td>
</tr>
<tr>
<td>Wannheden and Revenäs [33], 2020</td>
<td>Multimorbidities, known health condition or long-term conditions</td>
<td>4 (18)</td>
</tr>
<tr>
<td>Athilingam et al [34], 2017</td>
<td>Albert et al [35], 2018</td>
<td></td>
</tr>
<tr>
<td>Grossman et al [36], 2020</td>
<td>Macis et al [39], 2018</td>
<td></td>
</tr>
<tr>
<td>de Barros et al [32], 2013</td>
<td>Albinia and Hernandez [40], 2018</td>
<td></td>
</tr>
<tr>
<td>Hoffman et al [41], 2019</td>
<td>Alvarez et al [42], 2020</td>
<td></td>
</tr>
<tr>
<td>Oberschmidt et al [44], 2020</td>
<td>Pradhan et al [45], 2020</td>
<td></td>
</tr>
<tr>
<td>Oberschmidt et al [44], 2020</td>
<td>Vanoh et al [46], 2018</td>
<td></td>
</tr>
<tr>
<td>Du Preez and De La Harpe [47], 2019</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{a}AMD: age-related macular degeneration.

Recruitment Strategies

Table S3 (Multimedia Appendix 2) provides an overview of the recruitment strategies used and indicates studies in which no strategy was described. These include purposeful sampling, use of an outreaching approach (eg, attending local community support groups for the relevant study population), and invitation through partners. Moreover, the table provides an overview of the descriptions of the studied populations, including their characteristics and locations.

In 12 (55%) of the 22 studies, recruitment through partners was used to identify relevant and interested older people (for example, through patient associations) [26,28,31-33,37,38,41,42,44,45,47]. In 4 studies, purposeful sampling was used to recruit participants [35,36,40,46], and in 2 studies, outreach approaches were applied. These included contact made through local support groups for people with age-related macular degeneration (AMD) [27] and through posters and written materials physically placed in local public areas and sent out electronically (ie, through social media) [30]. Finally, in 4 studies, the recruitment strategy was not described [29,34,39,43].

The population groups in the studies were recruited based on specific parameters, including age, health conditions, and living conditions. These parameters can affect the ways in which people can be accessed and recruited. For instance, older people with variations of cognitive decline, such as dementia, are often perceived as difficult to access and are included in the development of new technology. However, in the included studies, older people with dementia and AMD were recruited using different methods, including outreaching (2/22, 9%) [27,30] and through partners (1/22, 5%) [31]. This demonstrates that the otherwise difficult-to-access population groups could be approached and recruited using appropriate approaches, such as an outreaching approach, meaning that a combination of methods of approaching and recruiting participants to target different groups of older people can ensure a broader representation in the design process of digital health technology. However, the most used recruitment method is invitation through partners, in which there is a risk of bias. For example, Oberschmidt et al [44] problematize recruiting through partners and highlight participant bias as a study limitation. The study emphasizes that the older people who participated were very active and outgoing. Thus, it is not representative of all older people. This shows a gap in the knowledge on the use of outreaching approaches to recruit older people with frailty and impairment in the design of digital health technologies.

Outcome of Involvement

Table S4 (Multimedia Appendix 3) provides a short description of the aims of the studies, as well as an overview of how the included populations contributed to the study outcome. The final column describes the health technologies developed in each study.
In 21 (95%) of the 22 studies, older people lived at home; in 1 (5%) study, the population consisted of inpatients at a hospital [42]. Thus, most of the technologies developed are aimed for people living at home. The technologies developed in these studies include various digital and web-based solutions, such as applications and digital platforms (18/22, 82%) [27-36,38,39,41,42,44-47], assistive technologies (2/22, 9%) [37,40], wearables (1/22, 5%) [26], and interactive caring television (1/22, 5%) [43]. In 64% (14/22) of studies, different variations in outcomes were presented, including those involving participant contributions; for example, how involvement led to a list of themes to be considered when developing an app based on end user needs [36] and how inputs from patients were used to identify design requirements for the interface [35] and to develop a platform [43]. Older people also assessed accessibility, leading to 7 features being included in an apps to prevent delirium in hospitalized older people [42].

Of the 22 studies, 8 (36%) studies did not report or reflect on the participants’ contributions; that is 5 (23%) studies did not report participants’ contributions [28,39,41,47] and 3 (14%) studies described the participants’ contribution in evaluating a prototype or by how they are involved and not by their contribution to the development of the technology [27,31,46]. Thus, gaps in the consistency of description of the value of involvement and outcome of the specific contribution of the participants were identified.

### Involvement Methods Used

Table S5 (Multimedia Appendix 4) provides an overview of the involvement methods used in the studies (eg, surveys and interviews). Moreover, it provides an overview of the time points when the methods were used to involve the participants in the design process, including needs identification, conceptualization, prototyping, or evaluation and further identifies whether the participants were included in one or several parts of the process.

The involvement of older people in this scoping review was assessed based on their involvement in 4 different parts of the development process. These four parts include the following:

1. needs identification, which is the first part of the development process in which end user needs are identified;
2. conceptualization, that is, the conceptualization of the final solution;
3. prototyping, that is, the development of a prototype; and
4. evaluation of the prototype.

In 9% (2/22) of studies, participants were included in all 4 parts of the development process including, needs identification, conceptualization, prototyping, and final evaluation [32,34]. In 23% (5/22) of studies, participants were included in 3 parts of the development process [33,41-43,45], and in 23% (5/22) of studies, older people were included in 2 parts of the development process [27,28,31,35,37]. This overview shows that 12 (55%) of the 22 studies present a combined ecosystem of methods, with consecutive steps that aim to ensure the involvement of older people in different parts of the development process of a digital health technology, from the identification of needs to the generation of ideas, cocreation of a specific product, and final evaluation.

In 45% (10/22) of studies, participants were included in 1 part of the development process. In 60% (6/10) of these studies, the involvement was in the final part of the process, that is, the evaluation of the prototype, using a variety of different involvement methods including, focus groups, workshops, feedback sessions, questionnaire, assessment of acceptance, usability assessment, and rating scales [26,29,30,38,39,46]. Thus, there is a gap in knowledge of the means to involve older people with frailty and impairment in all parts of the design process, including the initial needs assessment phase.

The most frequently used method to involve older people in the included studies (10/22, 45%) was interviews. Interviews were used for needs identification, prototyping, and evaluation. In addition, workshops and focus groups were also commonly reported and applied in all 4 parts of the development process. The participants were mostly involved in the final part of the development process, the evaluation (15/22, 68% of the studies), and in the initial needs identification (13/22, 59% of the studies), whereas participants were least involved in the prototyping process (9/22, 41% of the studies) and conceptualization phase (6/22, 27% of the studies).

Finally, in 14% (3/22) of studies, specific theories were used to inform the analysis [36,44,46]. Du Preez and De La Harpe [47] applied a grounded theory methodology through an iterative and simultaneous process of data collection, coding, category development, and data comparisons to understand the perceptions of older people regarding technologies to support aging in place. Greenhalgh et al [37] position their study within “critical ethnography,” referring to phenomenological philosophy touching upon Maurice Jean Jacques Merleau-Ponty and Martin Heidegger’s work on perception. Finally, Pradhan et al [45] used a constructivist grounded theory approach in their analysis. In total, 45% (10/22) of studies were conducted based on an existing framework or design concept (eg, feasibility study, scrum, PICTIVE [plastic interface for collaborative technology through video exploration] participatory design, and user-centered design framework) [26,27,31,32,35,36,38,39,42,46].

### Discussion

#### Overview

This scoping review sought to explore how older people with frailty and impairment are involved in various parts of the design processes of digital health technologies and to identify gaps or neglected steps in a user-involving design process. The focus has been on recruitment strategies, outcomes of involvement, and methods used to involve participants and address their perspectives, needs, and desires.

#### Principal Findings

In total, 3 gaps have been identified.

First, a gap in knowledge was identified regarding the use of different outreaching approaches to recruit older people with frailty and impairment in the design of digital health technologies. Involvement does not always begin during the recruitment process. Early involvement will enable an outreaching or alternative recruitment strategy to ensure a broad
representation of participants and access hard-to-reach populations. An outreaching approach was effectively used in 2 studies that recruited older people with dementia [30] and AMD [27]. However, the most used recruitment strategy in the current literature is through partners or by purposeful sampling. Second, a gap was identified in relation to the description of the value of involvement and outcome of the specific contribution of the participants. Reflection on and description of the outcomes of participants’ contributions is important. Our findings show that some studies successfully reflected the outcome of participants’ contributions.

In one-third (7/22, 32%) of the studies, the specific outcome of the contribution is not reflected upon, leaving a gap in understanding the degree and value of the involvement process. Third, a gap was identified in the knowledge regarding the means to involve participants in all parts of the design process, including the initial needs assessment phase. Using a variety of methods to involve older people with frailty and impairment in the design of new technologies is valuable, including focus groups, interviews, and workshops. An identified caution is the underrepresentation of involvement across the full design process as opposed to solely the final evaluation phase.

Involvement Starts With Recruitment

The findings indicate that choosing the right recruitment strategy is highly important to avoid recruitment bias and initiate a beneficial co-design process for older people with frailty and impairment. Therefore, reflecting on the use of different recruitment strategies is important to access a broader representation.

When recruiting participants, relevant factors should be considered, including how to reach the population of interest, as earlier studies have shown that older people and people with low digital skills are often left out or overlooked in the design process of new technologies. This lack of involvement can lead to increased inequity in health care services and a lack of access to new health technologies for those most in need [48].

The least commonly used recruitment strategy was the outreaching approach. The most used was purposeful sampling and invitation through partners. Oberschmidt et al [44] problematize recruiting through partners and highlight participant bias as a study limitation, emphasizing that the older people who participated were very active and outgoing. Future research need to focus on including older people who are less active and difficult to reach.

Hakobyan et al [27] benefited from using an outreaching approach to recruit people with AMD. The research group established contact with a support group for people with AMD. Over a period of 2 months, the research group attended 4 support group meetings to introduce themselves and learn more about their end users, including their capabilities and limitations. The research group found that the participants’ reluctance was sometimes related to their participation in research as an experimental subject, rather than an involved expert living with their specific condition. Together, the strategy to attend meetings for building relationships and obtaining a deeper insight into the reasons for the hesitation of potential participants ultimately enabled the research group to build a trusted relationship with the support group members, who eventually volunteered to participate in their study. Hassan et al [30] combined posters and written materials and distributed them physically and electronically (ie, via social media and email) to advertise the opportunity for involvement in the study. Using this method, approximately 25 people aged >65 years with dementia, memory problems, and mild cognitive impairments were recruited.

Finally, this scoping review found that some (4/22, 18%) studies that included older people with frailty and impairment had exclusion criteria that might have excluded relevant participants. These include cognitive, visual or hearing impairments, or severely limited dexterity in one or both hands [39], people with dementia [46], and those who required reading skills [38] or at least a secondary level of education (≥7 years) [46].

Description of Articulated Outcome of Involving Participants

The values of the involvement and contribution of the participants were explicitly addressed in 64% (14/22) studies. However, in 36% (8/22) studies, the contribution of involvement to the outcome was not described. This leaves a gap in the understanding of the degree and value of the involvement process.

The 14 studies addressed user involvement through a description of the involvement or reflection of the involvement. Athilingam et al [34] changed a prototype from being a chest-worn device used to monitor heart rate among patients with heart failure, to being a wrist-worn device, based on input from participants. In the study by Jacelon et al [38] the beta version of the user interface for “ASSISTwell,” a tablet app designed for older people to manage symptoms related to different chronic conditions, was developed using input from end users, retrieved through focus groups. De Barros et al [32] developed 4 apps for smartphone for the self-management of Parkinson disease, including (1) medication; (2) appointments; (3) my day, including disease status and symptoms; and (4) my data, including personal and health information, based on input retrieved through interviews, scoping sessions, focus groups, and usability testing with end users with Parkinson disease. Finally, Grossman et al [35] identified design requirements based on input from end users in the development of an interface to assist older people with heart failure.

In 36% (8/22) studies, the value of involvement and the contribution of the participants were not specified or reflected explicitly. Hakobyan et al [27] aimed through participatory and user-focused research to create a mobile assistive health care–related intervention for people with AMD to promote independent living. The methods used to involve the older people are described, including focus groups, observational studies, and design meetings, but the outcome of the involvement was not addressed. In other studies, participants were involved through surveys [40], interviews, and workshops [47] in the first part of the development process. When older people were involved in the initial and final aspects of design, this was often through interviews and observations [28], in 3 parts of the process including needs identification, prototyping, and evaluation of prototype [41], and in the final part of the
In only 9% (2/22) of studies, participants were included in all 4 stages of the design process. Athilingam et al [34] involved participants with heart failure in initial needs identification through needs assessment interviews. Moreover, the participants answered a questionnaire and provided input to the conceptualization and feedback on the design, features, and ease of use during the development phase. Finally, a feasibility study was conducted leading to significant changes in the software and design, which changed from a chest-worn device to a wrist-worn device. These elements were all a part of the study that focused on “patient engagement” with the purpose of achieving both a well-targeted solution for this specific population group and achieving persistent self-care and self-management, including positive health behavior for this group. De Barros et al [32] included participants through interviews and a scoping session with focus on daily routines, motivation mechanisms, medication-related behaviors, and specific requests for the smartphone app. Furthermore, focus groups and usability testing were conducted throughout the development of a smartphone app for self-management in people with Parkinson disease.

Among the identified methods of involvement, there was no indication that some were more successful than the others in identifying older people’s needs. As the purpose of this review was not to judge how and which kind of involvement method have been beneficial for the outcome of the studies, the findings highlight the methods that may be used to involve older people in design processes, so that their needs and desires are heard. Newell et al [51] stated that classic standards and guidelines for user-centered design are not always appropriate for including older people and people with disabilities. This suggests that “user-sensitive inclusive design” is a new way of including older people. This includes forming a close bond with the participants and using experimental techniques to involve older adults; for example, through theatrical techniques using actors instead of personas to impersonate a diverse group of older disabled adults.

There was no general difference in how and when participants with different conditions are involved. People with cognitive impairment were involved through focus groups, interviews, and workshops and were involved in all parts of the process in various studies. This could indicate that specific diseases are not limited to one specific involvement method and that there are several possibilities in relation to the involvement of older people with frailty and impairment in development processes. However, as several studies failed to report the outcome of involvement, it is impossible to draw strong conclusions about the appropriateness of the specific methods used. Earlier studies address the need for improving traditional methods used to involve end users and to consider limitations related to an aging population, including activities such as interviews, questionnaires, and observations. Moreover, earlier findings suggest that involvement and engagement in the initial steps of a development process increase the potential to create a technology that considers relevant limitations and characteristics related to older people [52].

For future work, consideration of the outcome in relation to the degree of involvement is relevant for further assessment. This
could include considering the level of involvement in relation to how many steps of the design process the end users are involved in and defining and assessing the degree of “active involvement” in using the different involvement methods. Further research on the effectiveness of these methods is required.

Recommendations from important stakeholders for engaging end users in the development of new medical and health technologies [20] have not been addressed in any of the included studies. There may be a need for specifications regarding how and in which steps user involvement and engagement should be performed. Our findings do not indicate a reason for excluding older people with frailty and impairment. It may be necessary for the regulatory bodies to clarify that these groups, if relevant to include, should not be excluded.

**Limitations**

Publications regarding digital health technology development are often conducted as part of the preparation for a certification process that is required by a funding body or in relation to a specific research goal. We may therefore have missed documentation in relation to commercial product development.

There may also be limitations in the representation of the specific disease addressed. The focus of this study was on older people with frailty and impairment, and thus, an overrepresentation of people with cognitive impairments is present, leaving out other major global disease burdens such as ischemic heart disease, diabetes, or multimorbidity, which may contribute to frailty.

Another limitation can be found in the definition of the applied technology used in this study. In this scoping review, studies involving welfare and mundane everyday technologies with no health-specific purpose were excluded, such as robot toilets, electrical curtains, robot vacuums, robot toilets, automated baths, and so on, as they are also part of a smart living environment and do not necessarily represent specific health technologies that were assessed in this study. Nevertheless, these studies may report the relevant methods on how to involve this population group for technology development, and the limitation to scope of the review is that we did not include those studies per the exclusion criteria.

**Conclusions**

This scoping review presents existing knowledge on how older people with frailty and impairment are involved in the design of digital health technologies that can contribute to their aging in place and also identifies gaps or neglected steps in a user-involving design process.

A gap in knowledge was identified regarding the use of outreaching approaches to recruit older people with frailty and impairment in the design of digital health technologies. The most commonly used recruitment strategy in the current literature is recruitment through partners or by purposeful sampling. The risk of bias in selecting participants is higher when using these forms of recruitment than when using an outreaching approach. However, it is important to emphasize that the literature does not suggest how the outcome of studies is affected by the different strategies.

Another gap was identified in the description of the value of involvement and the outcome of the specific contribution of the participants. Thus, reflection on the use of different involvement methods in future work could help evolve the existing practices and enable more older people, who are not commonly included in development processes, to take part in future projects.

Finally, a series of different methods used to involve older people in the development of digital health technologies was identified. However, a gap was identified in the knowledge regarding the means to involve the older people in all parts of the design process, including the initial needs assessment phase. The literature does not imply which part of the development process involvement is most beneficial. However, only few studies included participants throughout the development process, and an overrepresentation of participants involved at the end of the design process and underrepresentation of participants involved in the first steps of the design process were identified.

**Acknowledgments**

This research was funded by the European Union’s Horizon 2020 Research and Innovation Program under grant 101016848 and the Canadian Institutes of Health Research.

**Conflicts of Interest**

None declared.

Multimedia Appendix 1

Search strings.

[PDF File (Adobe PDF File), 49 KB - humanfactors_v10i1e37785_app1.pdf ]

Multimedia Appendix 2

Table S3. Recruitment strategies.

[PDF File (Adobe PDF File), 113 KB - humanfactors_v10i1e37785_app2.pdf ]
Multimedia Appendix 3
Table S4. Outcome of the study.

[PDF File (Adobe PDF File), 100 KB - humanfactors_v10i1e37785_app3.pdf]

Multimedia Appendix 4
Table S5. Involvement methods.

[PDF File (Adobe PDF File), 107 KB - humanfactors_v10i1e37785_app4.pdf]

References
17. Bergschöld J. Configuring dementia; how nursing students are taught to shape the sociopolitical role of gerontechnologies. Front Sociol 2018 Apr 9;3 [FREE Full text] [doi: 10.3389/fsoc.2018.00003]


25. Better systematic review management. Covidence. URL: https://www.covidence.org/ [accessed 2021-02-22]


Abbreviations

AMD: age-related macular degeneration  
EU: European Union  
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews  
SMILE: Smart Inclusive Living Environments  
WHO: World Health Organization

©Emilie Kauffeldt Wegener, Jenny M Bergschöld, Carly Whitmore, Marjolein Winters, Lars Kayser. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 27.01.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use,
Review

Exploring Patient Journey Mapping and the Learning Health System: Scoping Review

Amanda L Joseph¹,², BCom, MSc; Helen Monkman¹, BSc, MA, PhD; Andre Kushniruk¹, BA, BSc, MSc, PhD; Yuri Quintana¹,²,³,⁴, BSc, MSc, PhD

¹School of Health Information Science, University of Victoria, Victoria, BC, Canada
²Homewood Research Institute, Guelph, ON, Canada
³Division of Clinical Informatics, Beth Israel Deaconess Medical Center, Boston, MA, United States
⁴Harvard Medical School, Boston, MA, United States

Corresponding Author:
Amanda L Joseph, BCom, MSc
School of Health Information Science
University of Victoria
PO Box 1700 STN CSC
Victoria, BC, V8W 2Y2
Canada
Phone: 1 250 721 8575
Email: amandalynnjoseph@uvic.ca

Abstract

Background: Journey maps are visualization tools that can facilitate the diagrammatical representation of stakeholder groups by interest or function for comparative visual analysis. Therefore, journey maps can illustrate intersections and relationships between organizations and consumers using products or services. We propose that some synergies may exist between journey maps and the concept of a learning health system (LHS). The overarching goal of an LHS is to use health care data to inform clinical practice and improve service delivery processes and patient outcomes.

Objective: The purpose of this review was to assess the literature and establish a relationship between journey mapping techniques and LHSs. Specifically, in this study, we explored the current state of the literature to answer the following research questions: (1) Is there a relationship between journey mapping techniques and an LHS in the literature? (2) Is there a way to integrate the data from journey mapping activities into an LHS? (3) How can the data gleaned from journey map activities be used to inform an LHS?

Methods: A scoping review was conducted by querying the following electronic databases: Cochrane Database of Systematic Reviews (Ovid), IEEE Xplore, PubMed, Web of Science, Academic Search Complete (EBSCOhost), APA PsycInfo (EBSCOhost), CINAHL (EBSCOhost), and MEDLINE (EBSCOhost). Two researchers applied the inclusion criteria and assessed all articles by title and abstract in the first screen, using Covidence. Following this, a full-text review of included articles was done, with relevant data extracted, tabulated, and assessed thematically.

Results: The initial search yielded 694 studies. Of those, 179 duplicates were removed. Following this, 515 articles were excluded, as they did not meet the inclusion criteria. Next, 103 articles were read in full, and 95 were excluded, resulting in a final sample of 8 articles that satisfied the inclusion criteria. The article sample can be subsumed into 2 overarching themes: (1) the need to evolve service delivery models in health care, and (2) the potential value of using patient journey data in an LHS.

Conclusions: This scoping review demonstrated the gap in knowledge regarding integrating the data from journey mapping activities into an LHS. Our findings highlighted the importance of using the data from patient experiences to enrich an LHS and provide holistic care. To satisfy this gap, the authors intend to continue this investigation to establish the relationship between journey mapping and the concept of LHSs. This scoping review will serve as phase 1 of an investigative series. Phase 2 will entail the creation of a holistic framework to guide and streamline data integration from journey mapping activities into an LHS. Lastly, phase 3 will provide a proof of concept to demonstrate how patient journey mapping activities could be integrated into an LHS.

(JMIR Hum Factors 2023;10:e43966) doi:10.2196/43966
**Introduction**

What Is a Journey Map?

Journey maps are visualization techniques that can facilitate the diagrammatical representation of stakeholder groups by interest or function for comparative visual analysis [1,2]. Thus, in a health care context, journey maps can illustrate complex service delivery bottlenecks and describe the user experience across the continuum of care. There are 5 journey mapping techniques (Figure 1) that can each be used to illustrate a unique experience: (1) Mental (Cognitive) Model Map, (2) Experience Map, (3) Customer Journey Map, (4) Service Blueprint Map, and (5) Spatial Map [1-3]. Each mapping technique displays information distinctly and illustrates experiences in different contexts [1,2].

The benefit of these succinct visualizations lies in their ability to effectively illustrate intersections and relationships between organizations and consumers using products or services [4]. Therefore, journey maps can be used to help identify process pain points and highlight opportunities for improvement in various settings and contexts. Further, the visual findings of journey mapping activities can assist service providers and implementation scientists in effectively deploying resources to expand services or establish operational risks. As illustrated in Figure 1, the 5 journey mapping techniques have similarities and interrelationships yet provide distinct visual analyses [2]. Therefore, the sequence in which the mapping activities should be conducted depends on the intended outcome of the mapping exercise [2,5]. For example, the Mental (Cognitive) Model Map technique provides a visual analysis of the cognitive processes an individual may experience in their interactions with an activity, organization, or service [1-3]. The Experience Map technique displays the overall human experience of an individual’s activities not specific to an organization, product, or service [1-3]. Contrastedly, the Customer Journey Map technique illustrates a consumer’s interactions using a specific service, organization, or product [1-3]. Following this, the Service Blueprint Map technique illustrates experiences from a systems view [1-3] and relationships between organizational processes, individuals, and service delivery [1-3]. Lastly, the Spatial Map technique provides a broad view of relationships between processes, service delivery, and individuals [1-3].

**Figure 1.** Five journey mapping techniques adapted from previous studies [1-3].

What is a Learning Health System?

A learning health system (LHS) is a concept that emerged from the Institute of Medicine’s Roundtable on Evidence-Based Medicine [6]. The vision of an LHS is to “generate and apply the best evidence for the collaborative health care choices of each patient and provider; to drive the process of discovery as a natural outgrowth of patient care; and to ensure innovation, quality, safety, and value in health care” [6]. Further, Rubin and Friedman describe the LHS “as the tapestry that emerges from weaving together efforts across: health information management, health IT, patient engagement, clinical care, research, and public health arenas aimed at utilizing data, information, and knowledge to improve health” [7]. Since its introduction in...
2007, others have continued to adapt, redefine and expand on the concept and how it can be achieved. However, regardless of the varied definitions in the industry, the goals of an LHS are the same: “harness the power of data and analytics to learn from every patient, and feed the knowledge of ‘what works best’ back to clinicians, public health, and other stakeholders to create cycles of continuous improvement” [8].

The Continuous Knowledge Translation Loop of an LHS

An LHS can be conceptualized as a continuous learning microcosm that uses various data streams in the health care sector to improve service delivery and the human experience. As the health sector is multifaceted, there is a tremendous opportunity to more effectively use the often-fragmented data (ie, data stored in siloed and disparate health information systems) globally. An important aspect of an LHS lies in its potential to facilitate a continuous cycle of learning using health care data [8]. The strategic use of such data could allow external evidence from studies, reviews, and trials to inform practice and enrich the evidence base and, ultimately, the health system [9]. Further, the data, serving as a continuous feedback loop, could foster a mechanism in which evidence-based practices could be effectively used across the care continuum to catalyze systemic industry change. Specifically, the data gleaned from continuous data feeds could be aggregated and leveraged to improve service delivery in clinical practice and improve patient outcomes.

The Potential Value of Using Journey Map Data to Feed an LHS

As the health care sector operates on a 24/7 basis globally, an unquantifiable amount of data could be streamlined, examined, and used to improve efficiency in service delivery and holistically inform the health system. The fluid data cycle [8] outputs from each citizen (or patient), as they use various facets of the health system, could be captured and illustratively detailed via the 5 journey mapping techniques (Figure 1). Thus, the experiences of citizens and health care providers interfacing with the health system could be assessed and evaluated from multiple vantages and perspectives to inform the greater health ecosystem. Therefore, the data gleaned from the 5 journey mapping techniques [1-3] could provide a robust source and live data feed for a broader LHS and data repository. Additionally, integrating lived human experiences (ie, patient, physician, and caregiver journey mapping activities) into the design of health information systems (HIS) and health information technology (HIT) holds tremendous potential value for the creation of safer and more usable systems [10].

Objective

This paper aims to conduct a scoping review assessing the current state of the literature to establish a relationship between journey mapping techniques and LHSs.

Research Questions

- Is there a relationship between journey mapping techniques and an LHS in the literature?
- Is there a way to integrate the data from journey mapping activities into an LHS?
- How can the data gleaned from journey mapping activities be used to inform an LHS?

Methods

A scoping review, guided by the Arksey and O’Malley framework [11], was carried out by querying the following electronic databases: Cochrane Database of Systematic Reviews (Ovid), IEEE Xplore, PubMed, Web of Science, Academic Search Complete (EBSCOhost), APA PsycInfo (EBSCOhost), CINAHL (EBSCOhost), and MEDLINE (EBSCOhost). The key terms used were as follows: (Learning Health System) OR (Delivery of Healthcare), (Journey Mapping) OR (Patient OR Care) AND (Journey), and (Informatics) OR (Patient Health Information). The article evaluation began with a first screening in which 2 researchers independently assessed all articles by title and abstract using Covidence (Veritas Health Innovation), and articles were included (Figure 2) if they satisfied the following inclusion criteria:

- English articles with abstracts published between the years 2010 and 2022.
- Articles that referenced journey maps or mapping activities and an LHS.
- Articles that described user experiences in health care (eg, patients, caregivers, and physicians) and the LHS.

Subsequently, the 2 researchers independently screened and read the full-text articles to establish inclusion (Figure 2). Differences of opinion in article selection were resolved through discussion and team consensus. Lastly, the relevant data were extracted and tabulated for comparative analysis (Table 1), and the final selection of articles was assessed thematically to establish trends and themes in the literature.
**Figure 2.** Adaptation of PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) data flow diagram detailing article selection during the screening process [12].

![PRISMA Flow Diagram](image-url)

**Table 1.** Data extraction table illustrating the themes represented by each paper in this scoping review.

<table>
<thead>
<tr>
<th>Author</th>
<th>Study design</th>
<th>The need to evolve service delivery models in health care</th>
<th>The potential value of using patient journey data in an LHSa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Azar et al [13]</td>
<td>Descriptive</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Fung et al [14]</td>
<td>Pilot study</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Gartner et al [15]</td>
<td>Concept analysis and systematic review</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Sun et al [16]</td>
<td>Perspective</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Yu [17]</td>
<td>Editorial</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Joseph et al [1]</td>
<td>Scoping review</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Levine et al [18]</td>
<td>Pilot study</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Sharma et al [19]</td>
<td>Observational study</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>

aLHS: learning health system.  
b✓: denotes the themes represented in each paper.

**Results**

**Overview**

The initial search yielded 694 studies. Of those, 179 duplicates were removed in Covidence. Following this, a first screening of the article sample was conducted, and 515 articles were assessed (Figure 2). During the first screening phase, 412 articles were excluded, as they did not meet the inclusion criteria. Next, a full-text review of all 103 remaining articles was done. Of those, 95 were excluded, resulting in only 8 relevant articles that satisfied the inclusion criteria.

**Thematic Analysis**

After identifying relevant articles, each article was assessed thematically with data extracted and tabulated in Table 1. The findings from these articles can be subsumed into 2 overarching themes: (1) the need to evolve service delivery models in health care, which was expressed in 5 articles; and (2) the potential value of using patient journey data in an LHS, which was described in 3 articles. These 2 thematic categories will be examined in the subsequent sections.
Theme 1: The Need to Evolve Service Delivery Models in Health Care

With 5 articles stressing the urgency to evolve service delivery models in health care settings, it was the most prominent theme of the literature sample. In the article “The Indiana university center for healthcare innovation and implementation science: bridging healthcare research and delivery to build a learning healthcare system” [13], Azar et al detail that an “estimated 75,000 deaths every year could be prevented if high-quality care was more efficiently and effectively implemented” [13]. The authors quote the United States National Institute of Health, in that this considerable problem is not due to a paucity of knowledge, but rather poor incorporation of health care discoveries into daily practice [13]. Azar et al [13] clarify that over the past 3 decades, medical knowledge has increased, with 11 systematic reviews and 74 clinical trials being published every day, yet only 14% of these new findings are actually implemented in health care settings and translated into practice [13]. Therefore, to mitigate the risks to human health, the authors propose a paradigm shift in how health systems and service delivery should be conceptualized. They present 2 contrasting perspectives: (1) a traditional model of service delivery and (2) an innovative and adaptive model of health care service delivery. In the traditional model, organizations are viewed as machines that perform predictable, repeated tasks with replaceable parts that operate in stable and nondynamic settings [13]. In the adaptive model, health care systems are viewed as complex, dynamic, adaptive, and evolving systems comprised of a network of semiautonomous individuals (ie, health care professionals) who interact in nonlinear ways [13]. As health care needs and interactions are interdependent, interconnect, and changing over time [13], the authors insinuate that it is vital to design health care services to support the fluidity of systemic evolution. Thus, their article expressed the criticality of designing and developing an adaptable agile learning system that integrates hospital systems, population health, individual patients, and health care personnel [13].

Fung et al [14] present a systems approach to redesigning care in their article “Regional process redesign of lung cancer care: a learning health system pilot project.” Their novel approach enables timely access to cancer treatment for patients with lung cancer to a centralized specialty service that addresses clinical and operational challenges [14]. However, the authors caution that, despite its potential value, there is limited evidence of successful implementation of the LHS vision [14]. Thus, to streamline and operationalize the LHS concept, they developed the Ottawa Health Transformation Model as a regional approach to guide service delivery change and to integrate the nuances of the patient journey with best practices [14]. Further, the article laments that all facets of care need to be examined to address the complexity of health systems and to improve patient experiences holistically rather than just isolated parts [14]. The article concludes with the caveat that the value of the LHS approach in relation to service delivery is the creation of a system that can facilitate best practice adoption and fluid innovation [14]. Similarly, in their concept analysis and systematic review, Gartner et al [15] detail that a performant health care system is crucial for every country and that the current siloed health care business practices must be evaluated and challenged [15]. The authors suggest that fragmented health care services can compromise patient care, inhibit sustainable service delivery, and result in suboptimal use of financial and human resources [15]. Further, the authors state that repeated calls to improve the overall performance and quality of global health care delivery have occurred since 2001 [15]. The calls for transformational change in health care have been made by well-established national and international organizations such as the Institute of Medicine [15,20,21]; The National Academies of Sciences, Engineering, and Medicine [15,22]; and The World Health Organization [15,23,24]. Gartner et al [15] suggest that understanding the patient journey through an LHS view can facilitate the improvement of health care service delivery through a feedback loop in which data can be used to identify problem areas to support continuous improvement [15]. Lastly, in a similar yet contrasting view, Sun et al [16] express in their paper “Health management via telemedicine: learning from the COVID-19 experience” that telemedicine provides numerous opportunities to improve care efficiency, accessibility, and patient outcomes [16]. However, they state that many challenges exist, such as the digital divide, usability, and technology interoperability [16]. Further, the authors detail that the delivery of telemedicine services must evolve to support continuity of care throughout the patient journey [16]; specifically, by including the seamless integration of data from the clinical workflow of multidisciplinary care teams to support the LHS [16]. Nonetheless, they clarified that the implementation of a telemedicine business model must be supported by rigorous evidence-based practices, including clinical trials [16]. They warned that such precautionary measures are necessary to facilitate the seamless integration of telemedicine into routine care, ensuring the quality and safety of virtual care delivery [16]. Lastly, Yu et al [17] recount that data are only important and useful when they can be transformed into knowledge. In a health care context, the importance of data is realized when data sets of individual patients can be aggregated with similar patient data to inform patient populations [17]. Further, the value of clinical data lies in its interpretation in a clinical context among continuing care providers and when it is shared with the patient or their caregivers [17]. Additionally, the data set of a citizen (ie, patient) becomes of greater importance when it is combined with that of other citizens and when it can be aggregated for comparative statistical analysis to inform the health system on the health status of a population or subset [17].

Theme 2: The Potential Value of Using Patient Journey Data in an LHS

The potential value of using patient journey data in an LHS was expressed in 3 articles. In the article “Patient journey mapping: current practices, challenges and future opportunities in healthcare,” Joseph et al [1] describe how the data gleaned from patient journey maps could improve the health system by identifying varying patient experiences. Additionally, Joseph et al [1] detail that journey mapping approaches hold a significant value in improving complex health care processes for patients and providers alike. Further, the authors express that closely integrating patient journey mapping techniques into the health care system could create an LHS [1]. In their study
“Learning health system for breast cancer: pilot project experience,” Levine et al [18] report that clinicians need accurate and timely information on patient outcomes associated with various treatment modalities. Moreover, the authors describe that electronic health records are perceived to be helpful technologies, but access to patient data is often difficult [18]. However, despite the data accessibility challenges expressed in their study, the researchers were able to combine, read, and extract electronic health records data to view the patient journey [18]. Specifically, Levine et al [18] developed a prototype leveraging IBM Watson technology, with capabilities to validate artificial intelligence using natural language processing and to denote the clinical course of patients (ie, patient journey) in support of an LHS platform [18]. Their study findings illustrated a means by which the vision of an LHS could potentially be achieved by using artificial intelligence [18]. Despite the preliminary nature of their study, the authors were able to demonstrate that the hospital had the necessary data to formulate a view of the patient journey, which could be extracted and used in ways to support clinical decision-making [18]. Lastly, in their observational study, Sharma et al [19] used an incremental and iterative approach, engaging administrative and clinical domain experts to demonstrate that human actors, rather than IT, are the central focus of data movement [19]. The authors evaluated a kidney transplant referral pathway and established the relationship between human actors, organizations, the complexity of data administration, and data flow bottlenecks [19]. Their study illustrated the manual and often cumbersome tasks that clinical staff must perform to access and visualize health data from fragmented IT systems [19]. The authors express broadly that IT systems that are not interoperable can lead to data access challenges and complicate the clinical workflow and health care providers’ ability to effectively and efficiently perform their job functions [19]. They further reveal that in a kidney transplant referral context the lack of centralized and timely access to patient data can delay patients’ registration on the transplant list, as the time and effort to complete referral forms are greatly increased [19]. Sharma et al [19] propose that an LHS with linked patient data can improve population health outcomes and inform interventions by providing timely and intuitive access to health information.

Summary of Findings

Despite the comprehensive search, the research questions were only partly satisfied. The first research question, “Is there a relationship between journey mapping techniques and an LHS in the literature?” was demonstrated in both thematic categories. There is a relationship and a need for an innovative approach to health care design and service delivery. As shown in Table 1, five articles exemplify the need to evolve service delivery models in various scenarios in health care. Three articles provide insight into the potential value of using patient journey data to inform an LHS. The second research question, “Is there a way to integrate the data from journey mapping activities into an LHS?” was not comprehensively addressed, and an actionable, scalable plan was not provided in the literature. The third question, “How can the data gleaned from journey mapping activities be used to inform an LHS?” was satisfied by the scoping review findings (Table 1). Many articles provided examples of operational gaps and scenarios in which patient care could be compromised due to a lack of timely, interoperable, and accessible data.

Discussion

Principal Findings

This study has presented a scoping review using articles from the following electronic databases: Cochrane Database of Systematic Reviews (Ovid), IEEE Xplore, PubMed, Web of Science, Academic Search Complete (EBSCOhost), APA PsycInfo (EBSCOhost), CINAHL (EBSCOhost), and MEDLINE (EBSCOhost). As evidenced by the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram (Figure 2), of 694 initially screened articles, only 8 satisfied the inclusion criteria. Within the articles that met the inclusion criteria, we identified 2 important themes: (1) five articles stressed the need to evolve service delivery models in health care, and (2) three articles described the potential value of using patient journey data in an LHS. Despite the robust search strategy and databases used, there was a dearth of literature discussing a relationship between journey mapping and LHSs. Therefore, the first research question, “Is there a relationship between the journey mapping techniques and an LHS in the literature?” was only partially satisfied. Despite the included articles [1,13-19] providing various scenarios and applications of the relationship potential and how journey mapping could support value-based and patient-centric care strategies for LHSs [25], more research is required in this arena. Further, given the current state of the literature, we could not address the second research question, “Is there a way to integrate the data from journey mapping activities into an LHS?” Although the urgency of timely access to centralized, high-quality, interoperable data was prominent in the literature, a comprehensive road map or framework was not provided to integrate the data specifically from journey mapping activities into an LHS. Lastly, the third question, “How can the data gleaned from journey map activities be used to inform an LHS?” was satisfied by the scoping review findings (Table 1). Many articles provided examples of operational gaps and scenarios in which patient care could be negatively impacted by workflow bottlenecks or disruptive technologies. However, specifically how patient journey map data could be used to inform the continuous learning feedback loop of an LHS, which could inform evidence-based practices, was not provided. Further, the article sample did not provide detail on how the 5 journey mapping techniques (Figure 1) could independently or collectively provide rich and diverse continuous data supply (ie, a continuous knowledge translation loop) for an LHS.

To address the shortcomings in the literature, the authors will continue this line of investigation to establish a relationship between the concept of an LHS and the 5 journey mapping techniques: (1) Mental (Cognitive) Model Map, (2) Experience Map, (3) Customer Journey Map, (4) Service Blueprint Map, and (5) Spatial Map [1-3]. Specifically, this scoping review will be phase 1 of an investigative series. Phase 2 will entail the creation of a holistic framework to guide and streamline data integration from journey mapping activities outputs into an
LHS. Lastly, phase 3 will provide a proof of concept to demonstrate how journey mapping activities could be integrated into an LHS.

Limitations
As this is a preliminary study, the scoping review was limited to only digital articles in English; therefore, other relevant articles could have been omitted based on the study design. Future studies could include paper-based searches and searches in other languages. Moreover, journey maps are not widely or consistently used in the health care sector [1,2], which may have contributed to the study’s small sample of only 8 relevant articles. Similarly, although poised to improve health care sustainably through smart and efficient data use, LHS is a relatively novel and emerging concept in the health care landscape [6].

Conclusions
This paper expressed the criticality and urgent need of global health care transformation to support the sustainable delivery of health care services. Additionally, it was revealed that current health systems are not adequately using the health data in which they aggregate institutionally. Consequently, fragmented and siloed data are stored in disparate HIS and HITs on a global scale. Thus, there is a dire need to design and develop an agile and interoperable LHS that can integrate global data from health care organizations, populations (ie, citizens, patients, caregivers, physicians, and health care stakeholders), HIS, and HIT. Journey mapping activities provide an opportunity and a conduit to streamline data into uniform and usable formats. Thus, the knowledge gap related to integrating the data from journey mapping activities into an LHS highlighted the importance of using the data from patient experiences to enrich an LHS and provide holistic care. Moreover, the journey mapping visualizations of the 5 mapping techniques (Figure 1) could identify operational issues such as staffing shortages, clinical workflow bottlenecks, and other factors that could negatively impact patient care [1,2]. The visualizations could also illustrate scenarios where health care service design and delivery could be stifled or affected from a clinical lens by physician burnout and cognitive impairment from alert fatigue [26]. Integrating the data from the 5 journey map techniques [1-3] into an LHS promises to improve health care service delivery and patient outcomes by providing a continuous supply of data to support patient-centric health care solutions that meet the goals of patients and providers.

Acknowledgments
ALJ received funding from the Natural Science and Engineering Research Council of Canada Visual and Automated Disease Analytics Program.

Authors’ Contributions
ALJ conceptualized the study design, conducted the literature search and data analysis, and created the first draft. HM assisted with the literature search and data analysis. AK and YQ contributed to the refinement of the paper. ALJ led the writing, with all the other co-authors commenting on subsequent drafts. All authors approved the publication of the final version.

Conflicts of Interest
AK is editor in chief of JMIR Human Factors. YQ is an editorial board member of JMIR Human Factors. The other authors declare no conflicts of interest.

References


Abbreviations

   HIS: health information systems
   HIT: health information technology
   LHS: learning health system
   PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
© Amanda L Joseph, Helen Monkman, Andre Kushniruk, Yuri Quintana. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 27.02.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Assessing the Quality and Impact of eHealth Tools: Systematic Literature Review and Narrative Synthesis

Christine Jacob¹, PhD; Johan Lindeque², PhD; Alexander Klein³, MSc; Chris Ivory⁴*, PhD, Prof Dr; Sabina Heuss²*, PhD, Prof Dr; Marc K Peter²*, DBA, Prof Dr

¹FHNW - University of Applied Sciences Northwestern Switzerland, Windisch, Switzerland
²FHNW - University of Applied Sciences Northwestern Switzerland, Olten, Switzerland
³Medical Affairs (Personalised Healthcare and Patient Access), F Hoffmann-La Roche Ltd, Basel, Switzerland
⁴Innovation Management, Mälardalens University, Västerås, Sweden
*these authors contributed equally

Corresponding Author:
Christine Jacob, PhD
FHNW - University of Applied Sciences Northwestern Switzerland
Bahnhofstrasse 6
Windisch, 5210
Switzerland
Phone: 41 798769376
Email: christine.k.jacob@gmail.com

Abstract

Background: Technological advancements have opened the path for many technology providers to easily develop and introduce eHealth tools to the public. The use of these tools is increasingly recognized as a critical quality driver in health care; however, choosing a quality tool from the myriad of tools available for a specific health need does not come without challenges.

Objective: This review aimed to systematically investigate the literature to understand the different approaches and criteria used to assess the quality and impact of eHealth tools by considering sociotechnical factors (from technical, social, and organizational perspectives).

Methods: A structured search was completed following the participants, intervention, comparators, and outcomes framework. We searched the PubMed, Cochrane, Web of Science, Scopus, and ProQuest databases for studies published between January 2012 and January 2022 in English, which yielded 675 results, of which 40 (5.9%) studies met the inclusion criteria. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines and the Cochrane Handbook for Systematic Reviews of Interventions were followed to ensure a systematic process. Extracted data were analyzed using NVivo (QSR International), with a thematic analysis and narrative synthesis of emergent themes.

Results: Similar measures from the different papers, frameworks, and initiatives were aggregated into 36 unique criteria grouped into 13 clusters. Using the sociotechnical approach, we classified the relevant criteria into technical, social, and organizational assessment criteria. Technical assessment criteria were grouped into 5 clusters: technical aspects, functionality, content, data management, and design. Social assessment criteria were grouped into 4 clusters: human centricity, health outcomes, visible popularity metrics, and social aspects. Organizational assessment criteria were grouped into 4 clusters: sustainability and scalability, health care organization, health care context, and developer.

Conclusions: This review builds on the growing body of research that investigates the criteria used to assess the quality and impact of eHealth tools and highlights the complexity and challenges facing these initiatives. It demonstrates that there is no single framework that is used uniformly to assess the quality and impact of eHealth tools. It also highlights the need for a more comprehensive approach that balances the social, organizational, and technical assessment criteria in a way that reflects the complexity and interdependence of the health care ecosystem and is aligned with the factors affecting users’ adoption to ensure uptake and adherence in the long term.

(JMIR Hum Factors 2023;10:e45143) doi:10.2196/45143

KEYWORDS
eHealth; mobile health; mHealth; digital health; technology assessment; technology adoption; technology implementation
Introduction

Background

Research has shown that eHealth solutions may help optimize the quality of health care services [1-6] but also that the lack of a standardized assessment approach makes it challenging to select the appropriate tool for a particular purpose in a particular context [7-9]. eHealth tools continue to grow in number, creating a cluttered landscape that can be hard to navigate. Regarding mobile health apps alone, there are >300,000 available in the app stores, and >200 new apps are added daily [10]. Stakeholders, including patients, clinicians, payers, and other industry players such as pharmaceutical companies, face challenges in identifying quality in this crowded space [7,8]. It has also been established that users are faced with a situation where only a fraction of the available solutions are in fact appropriate for use [11], with considerable variation in the evidence supporting the different eHealth interventions [12]. Hence, there is a need for standardized assessment criteria to support informed decision-making with respect to eHealth tool evaluation [8].

Technological advancements have opened a path for many technology providers to easily develop and introduce eHealth tools to the public. The use of these tools is increasingly recognized as a critical quality driver in health care [13]; however, choosing a quality tool from the myriad of tools available for a specific health purpose is challenging. Moreover, rapid technological development means that many eHealth tools remain unevaluated by researchers [9,14], leaving potential users largely uninformed about their quality, veracity, safety, and fit [15]. Owing to this lack of proper assessment mechanisms, previous researchers that tried to assess existing apps have concluded that many eHealth tools that hit the market lack some relevant functionality and features [16] or do not fully satisfy users’ needs [17]. Furthermore, the crowded eHealth landscape compared with the number of approved prescription drugs, for instance, makes it quite challenging for both clinicians and patients to find, evaluate, and adopt the right eHealth tools [18]. Quite often, clinicians find themselves in a situation where they do not know which tool to use or recommend [19,20]. Failure to properly assess criteria such as the accuracy and appropriateness of eHealth tools can also compromise patient safety [21]. Ultimately, the lack of standardized and rigorous assessment frameworks results in tools that do not always meet high-quality standards across multiple domains [17].

Objectives

The aim of this study was to build a better understanding of the different criteria used to assess the quality and impact of eHealth technologies. We adopted the World Health Organization (WHO) definition of eHealth as “the cost-effective and secure use of information and communications technologies in support of health and health-related fields, including health care services, health surveillance, health literature, and health education, knowledge and research” [22]. Furthermore, this review focused on patient-facing eHealth tools, including self-management tools and remote eHealth solutions, rather than tools used within and between care providers (eg, health care professional videoconferences or electronic health record integration) or health data analytics systems used at the population level.

Accordingly, a systematic review was conducted to provide a precise and up-to-date description of the different criteria used in published research to assess the quality and impact of eHealth tools from technological, social, and organizational perspectives. It also reflected on the potential implications and suggested directions for relevant stakeholders on how to best assess the eHealth tools that they are considering. This work builds on and expands the initial findings of a previous research project that investigated the sociotechnical factors affecting mobile health adoption from patients’ and clinicians’ perspectives, which have already been published [23,24].

Findings from this study will help inform clinicians, pharmaceutical executives, insurance professionals, technology providers, and policy makers by presenting them with an up-to-date and comprehensive review of the different criteria used to assess the quality and impact of eHealth tools as reported in the academic literature. This can guide them in making more informed decisions about which tools to use, endore patients, invest in, partner with, or reimburse based on their potential quality and impact.

Methods

Overview

The methods for this review were drawn from the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [25] and the Cochrane Handbook for Systematic Reviews of Interventions [26], both of which provide guidance toward a rigorous and reliable literature review methodology. The review methods were defined in advance, and the protocol was published in the Research Registry (reference: reviewregistry1291) and is available on the web to promote transparency [27]. This analysis did not require any major divergence from the initial protocol. The research question that guided this review was as follows: what are the technical, social, and organizational criteria that must be considered when assessing the quality and impact of eHealth tools?

Search Strategy

A search of the PubMed, Cochrane, Web of Science, Scopus, and ProQuest databases in January 2022 identified relevant studies. The scope of this review was narrowed to studies published in English between January 2012 and January 2022. Only original, peer-reviewed, and published papers were included in this study. Other forms, such as editorials, unsystematic reviews, interviews, commentaries, unstructured observations, and position papers, were excluded. We decided not to include articles based on manual searches of reference lists in alignment with the guidance of the Cochrane Handbook for Systematic Reviews of Interventions that “positive studies are more likely to be cited” and “retrieving literature by scanning reference lists may thus produce a biased sample of studies” [26].

The search string shown in Textbox 1 was developed according to the participants, intervention, comparators, and outcomes framework. The authors limited the search of this search string.
to the manuscript title to make sure that the resulting papers were about eHealth assessment criteria as a whole, not individual assessments of pilot studies singling out specific tools. Comparators were not applicable to this study.

Textbox 1. The search string according to the participants, intervention, comparators, and outcomes framework.

<table>
<thead>
<tr>
<th>Participants: patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Focus on patient-facing eHealth technologies, including self-management tools and remote eHealth solutions, rather than tools used within and between care providers (eg, health care professional videoconferences or electronic health record integration) or health data analytics systems used at the population level</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Intervention: eHealth</th>
</tr>
</thead>
<tbody>
<tr>
<td>“eHealth” OR “mobile health” OR “Telehealth” OR “mHealth” OR “mobile applications” OR “mobile apps” OR “telemonitoring” OR “app” OR “online health apps” OR “digital health” OR “health apps” OR “health platforms”</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outcome: assessment criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>AND (“assessment” OR “assess” OR “evaluation evaluating” OR “validation” OR “impact” OR “effectiveness” OR “efficacy” OR “quality”)</td>
</tr>
<tr>
<td>AND (“criteria” OR “framework” OR “method” OR “methodology” OR “methodologies” OR “measurement” OR “toolkit” OR “tool” OR “tools” OR “approach” OR “scorecard” OR “path”)</td>
</tr>
</tbody>
</table>

Study Selection

In total, 2 researchers (CJ and JL) were involved in the screening, eligibility, and inclusion phases, and any divergence was agreed upon through discussion between them. In cases where they could not reach an agreement, a third reviewer (SH for social or health-related criteria, CI for organizational criteria, and MP for technical criteria) discussed it with them and made the final decision. The practice partner (AK) ensured that the naming and categorization of the assessment criteria were relevant and meaningful from a practice point of view. The research team used the open-source app Rayyan (Qatar Computing Research Institute) to facilitate collaborative screening by the team [28]. Screening lasted from February 2022 to June 2022. The inclusion and exclusion criteria are detailed in Textbox 2 and were developed according to the participants, intervention, comparators, and outcomes framework.

After completing screening and resolving any conflicting views among the researchers, the selected full texts were assessed for eligibility independently by CJ and JL. Any disagreements were resolved through discussion with SH for social or health-related criteria, CI for organizational criteria, and MP for technical criteria. The risk of bias was assessed using the Critical Appraisal Skills Programme (CASP) checklist [29]. The checklist is provided in Multimedia Appendix 1, and it evaluates the following key quality criteria of the included studies: whether there was a clear statement of the aims of the research, whether the methodology was appropriate for the research objectives, whether the research design was appropriate to address the aims, whether the recruitment strategy was appropriate for the aims of the research, whether the data were collected in a way that addressed the research issue, whether the role of the researchers was adequately considered, whether ethical issues were considered, whether the data analysis was sufficiently rigorous, whether there was a clear statement of findings, and whether the researchers discussed the contribution the study made to existing knowledge or understanding (eg, did they consider the findings in relation to current practice or policy or relevant research-based literature). A Microsoft Excel (Microsoft Corp) sheet with the results of the appraisal of the included studies can be accessed in Multimedia Appendix 2 [15-21,30-62].
Textbox 2. Inclusion and exclusion criteria according to the participants, intervention, comparators, and outcomes framework.

### Inclusion criteria
- **Participants**
  - Focused on patients
- **Intervention**
  - Focused on patient-facing eHealth tools, including self-management tools and remote eHealth solutions
- **Comparators**
  - Does not apply
- **Outcomes**
  - Addresses the different criteria used to assess the quality and impact of eHealth tools regardless of the condition
- **Publication type**
  - Original, peer-reviewed, and published papers
- **Time frame**
  - Studies published between January 2012 and January 2022
- **Language**
  - Studies published in English

### Exclusion criteria
- **Participants**
  - Focused solely on clinicians or technology providers
- **Intervention**
  - Tools used within and between care providers (e.g., health care professional videoconferences or electronic health record integration) or health data analytics systems used at the population level
- **Comparators**
  - Does not apply
- **Outcomes**
  - Individual assessments of pilot studies singling out specific tools
- **Publication type**
  - Editorialis, interviews, commentaries, unstructured observations, and position papers
- **Time frame**
  - Studies published before January 2012 or after January 2022
- **Language**
  - Studies published in other languages

### Data Collection and Synthesis
The variety of procedures and results that were identified in the included studies was not homogeneous enough to enable a quantitative analysis of the data. Therefore, a narrative synthesis was used and structured around the social, organizational, and technical criteria used to assess the quality and impact of eHealth tools. NVivo (QSR International), a computer-assisted qualitative data analysis software, was used to assist with this task.

Data coding began with a preliminary data extraction grid that included themes based on previous research and technology acceptance frameworks; the initial codebook was informed by our previous work that aggregated the factors affecting adoption from patients’ and clinicians’ perspectives [23,24,63]. More
codes were added as they emerged during the review process. Thematic analysis by Braun and Clarke [64-66] was used to identify and extract themes under the social, technical, and organizational assessment criteria addressed in the research question. Social criteria included any social-related elements, such as the effects of people and groups influencing one another through culture; technical criteria included elements related to the material sides of the technology, such as its ease of use and usability; and organizational criteria were linked to elements such as resources and workflow. The phases of the thematic analysis are explained in detail in Multimedia Appendix 3. The 7 key phases were data familiarization; initial code generation; searching for themes; reviewing themes; defining and naming themes; linking themes to explanatory frameworks; and, finally, producing the report. This process lasted from June 2022 to September 2022.

**Theoretical Framework**

Health care technologies are generally more complex than tools that address a specific user need. They typically serve patients with comorbidities who are mostly treated by multidisciplinary teams of clinicians potentially working across more than one organization. This particular nature of the health care sector calls for a wider view that goes beyond a tool’s technical aspects as health care technology cannot be successfully implemented in isolation from the broader context in which it is being used [63]. Therefore, the authors were guided in their thinking by the sociotechnical theory, which has at its core the idea that the design and performance of any innovation can only be understood and improved if both “social” and “technical” aspects are brought together and treated as interdependent parts of a complex system [67]. In social studies of technology and, more specifically, the sociotechnical theory, technology, roles, and practices and organizational structures are viewed as interacting parts of mutually interdependent collections of elements [67]. This position is aligned with what several scholars have recommended (explaining that many of the broadly used frameworks adopt a technology-centered view focusing on the technological aspects [68,69]): a shift to multidimensional models that go past technology to encompass the surrounding context as well as societal and implementation factors [68-71]. Therefore, the resulting criteria go beyond the technical quality of eHealth tools to also cover all other relevant aspects, such as social and organizational criteria.

**Results**

**Study Selection Flow and Characteristics of the Included Studies**

The PRISMA study selection flow diagram in Figure 1 depicts the flow of information through the different phases of the systematic review. It maps out the number of records identified, included, and excluded and the reasons for exclusion. This process resulted in the inclusion of 40 articles for the qualitative synthesis [15-21,30-62]. Multimedia Appendix 4 [15-21,30-62] presents the sample characteristics of the included studies from research methodology, geographical, and clinical focus perspectives.

![Study selection flow diagram based on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines.](https://humanfactors.jmir.org/2023/1/e45143)

**Critical Appraisal**

We assessed the quality of the included studies using the CASP checklist for qualitative studies [29]. We chose the CASP because of the diversity of methodologies used in the included studies and the narrative nature of our own synthesis (as opposed to meta-analysis and more quantitative methodologies) and because it is the most commonly used tool for quality appraisal in health-related qualitative evidence synthesis, with endorsement from the Cochrane Qualitative and Implementation
Methods Group [72]. The included studies encompassed diverse methodologies, including quantitative, qualitative, and mixed methods as well as systematic literature reviews; hence, some of the questions on the checklist were not applicable to all types of studies. Scores were not assigned as this was not recommended by the checklist [29].

On the basis of the critical appraisal, of the 40 studies, 4 (10%) did not clearly justify their choice of study design but still used a design that was suitable for their objectives, 3 (8%) did not provide sufficient details on the profiles of the assessors and implications for potential bias, 5 (12%) did not report whether the study procedure was reviewed for ethics approval or how they protected the privacy of the participants, 12 (30%) were not clear enough about their data analysis strategy and whether it was sufficiently rigorous, and 4 (10%) did not sufficiently discuss the practical or policy implications of their findings. The quality assessment results are provided in Multimedia Appendix 2.

Studies were not excluded based on quality assessment outcome as this was unlikely to have a major influence on the definition of the assessment criteria and the resulting aggregated framework. However, the assessment provided a general idea of the quality of the development processes of the existing frameworks and, therefore, the strength of the evidence [73]. This will be discussed in more detail in the Discussion section when addressing the challenges with existing initiatives and frameworks.

Frameworks and Guidelines That Resulted From or Were Used in the Included Studies

Several publications (21/40, 52%) did not mention the use of a framework; however, there were 19 different frameworks or guidelines used, and 22% (9/40) of the studies resulted in the creation of a new assessment framework. Figure 2 presents the diversity of the frameworks used in or resulting from the included studies according to their occurrence. A framework resulting from a study means that this framework was the end result of the work in that study, whereas a framework used in a study was the starting point rather than the outcome of that study.

Stoyanov et al [55] created the Mobile App Rating Scale (MARS), and Roberts et al [21] adapted it, creating the adapted MARS (A-MARS) to make it appropriate for the evaluation of both mobile phone apps and e-tools, whereas EVALAPPSS was the outcome of the work by Robles et al [62]. The Clinical Information Quality (CLIQ) framework for digital health resulted from the work by Fadahunsi et al [37], whereas the work by Baumel et al [32] resulted in the creation of Enlight, a comprehensive quality and therapeutic potential evaluation tool for mobile and web-based eHealth interventions.

Garell et al [38] focused on evaluating digital health services according to current legislation by creating a framework for assessing the legal challenges in Digital Health (LCDH) framework, whereas the Medication Adherence App Quality (MedAd-AppQ) assessment tool resulted from the work by Ali et al [16]. The updated National Institute for Health and Care Excellence Evidence Standards Framework (NICE-ESF) for digital health and care technologies was the result of the work by Unsworth et al [56], whereas Varshney et al [57] created the Review, Assess, Classify, and Evaluate (RACE) process, and Camacho et al [18] created the Technology Evaluation and Assessment Criteria for Health Apps (TEACH-apps) process.

Of the frameworks and guidelines that were included in the studies, only 2 were used twice, and the rest were only used once. The Health Care Information and Management Systems Society criteria framework [74,75] was used by Stoyanov et al [55] and Wildenbos et al [61]. The reach, effectiveness, adoption, implementation, and maintenance (RE-AIM) framework [76] was used by Blackman et al [34] and de La Vega et al [35], whereas the American Psychiatric Association (APA) app evaluation framework [77] was used by Camacho et al [18]. The App Quality Evaluation (AQEL) framework [78] was used by DiFilippo et al [36], and the Behavior Interventions Using Technology (BIT) framework [79] was used by de La Vega et al [35].

The Consensus-Based Standards for the Selection of Health Measurement Instruments initiative [80,81] was used by Muro-Culebras et al [50], whereas the Digital Health Scorecard [8,82] was used by Sedhom et al [17], and the European Network for Health Technology Assessment (EUNetHTA) Core Model [83,84] was used by von Huben et al [60]. Stoyanov et al [55] used the European Union UsabilityNet [85] and the Nielsen Norman user experience criteria [86]. The Food and Drug Administration (FDA) precertification program [87,88] was used by Alon et al [15], whereas Ali et al [16] used a version of the Health On the Net Foundation code of conduct [89,90] that was adapted to assess the reliability and credibility of medical apps [91,92]. The quality improvement framework of the Institute of Medicine (IOM) [93] was used by Lee et al [46].

The Intervention Scalability Assessment Tool (ISAT) [94] was used by Azevedo et al [30], whereas the National Institute for Health and Care Excellence behavior change guidance (NICE BCG) [95] was used by McMillan et al [48], and the Prioritering efter Nyttogrunder (PENG; translated as “Prioritizing based on contribution of benefits”) evaluation tool [96] was used by Parv et al [52]. Finally, the Replicating Effective Programs (REP) framework [97] was used by Camacho et al [18]. Multimedia Appendix 5 [8,15-18,21,30,32,34-38,46,48,50,52,55-57,60-62,74-93,95-97] presents the frameworks and guidelines that resulted from or were used in the included studies and provides more details on their contexts and the assessment criteria that each of them encompassed.

**Synthesized Assessment Criteria**

We synthesized similar measures from the different papers, frameworks, and initiatives, resulting in 36 unique criteria that mirrored all the relevant assessment methods that were cited in the included papers. It is worth noting that some of the criteria may fit into more than one category but were placed in the best-fitting category because of their importance and impact. For example, inclusive design could be considered a design aspect and could have been included in the design cluster under the technical assessment criteria; however, given its importance for human centricity and its social implications for health care equity, it was placed in that cluster instead. We also deliberately included assessment criteria that apply to high-risk eHealth tools as it allowed us to identify a more extensive list of criteria with the expectation that not all criteria will necessarily apply to lower-risk eHealth tools. For instance, the patient safety assessment criteria mostly apply for high-risk tools and would be less relevant for low-risk tools that do not endanger patient safety.

Using sociotechnical theory as a guide, we classified the relevant criteria into technical, social, and organizational criteria, as detailed in Figure 3, which shows the aggregated criteria from all the included studies, the frameworks that mentioned each criteria listed in brackets, and their occurrence. The double-ended arrows in the figure signal the interplay between the technical, social, and organizational aspects. For instance, the social criteria related to human centricity and inclusive design would also affect and be affected by the technical criteria related to the tool’s design, such as usability. Similarly, the health care organization organizational criteria, such as infrastructure and implementation, will affect and be affected by the technical criteria related to data integration and interoperability. Multimedia Appendix 6 [15-21,30-62] reflects the assessment criteria classified according to the sociotechnical approach, the respective frameworks where they prevailed, their occurrences in the included studies, their definitions, and the respective references.

Technical Assessment Criteria

The technical assessment criteria were grouped into 5 clusters: technical aspects, functionality, content, data management, and design. The technical aspects cluster includes technical reliability and stability (BIT, MARS, A-MARS, NICE ESF, EUEnHTA; n=14/40, 35%), which typically refer to the system quality of the tool from a technical perspective and potential technical issues (eg, errors, freezing, and response time of the application); training and documentation (REP, TEACH-apps, NICE BCG, A-MARS, EUEnHTA, EVALAPPS; n=18/40, 45%); support and help resources (REP, TEACH-apps, A-MARS; n=14/40, 35%); and feedback mechanisms (REP, TEACH-apps, A-MARS; n=2/40, 5%).

Functionality

- Feature definition, attributes, functionality, purpose, and user requirements (NICE BCG, RACE, TEACH-apps, A-MARS, A-MARS, EUEnHTA, EVALAPPS; n=18/40, 45%)
- Feature usefulness, utility, and relevance (MedAd-AppQ, Enlight, A-MARS, NICE ESF, EVALAPPS; n=15/40, 38%)
- Feature convenience (MedAd-AppQ; n=3/40, 8%)

Content

- Content quality (Enlight, A-MARS, NICE ESF, EUEnHTA, EVALAPPS; n=15/40, 38%)
- Credibility (Enlight, APA, CLIQ, MARS, A-MARS, NICE ESF, EUEnHTA, EVALAPPS; n=15/40, 38%)
- Validity and reliability (MedAd-AppQ, NICE ESF, EVALAPPS; n=10/40, 25%)

Data management

- Privacy and security (Pre-Cert, MedAd-AppQ, Enlight, APA, CLIQ, LCDH, NICE BCG, MARS, NICE ESF, EUEnHTA, EVALAPPS; n=26/40, 65%)
- Data integration and interoperability (APA; n=7/40, 18%)

Design

- Usability (Enlight, APA, BIT, A-MARS, CLIQ, MARS, NICE ESF, EUEnHTA, EVALAPPS; n=27/40, 68%)
- Visual design (Enlight, MARS; n=12/40, 30%)
- Timeliness (OM, A-MARS; n=4/40, 10%)

Human centrality

- User engagement, customization, tailoring and user control (Enlight, REP, TEACH-apps, NICE BCG, MARS, A-MARS, RACE; n=17/40, 43%)
- Behavioural change and persuasiveness (Enlight, A-MARS, NICE BCG, NICE ESF, RACE; n=14/40, 35%)
- Equity, accessibility and inclusiveness (ION, MARS, A-MARS, NICE ESF, EUEnHTA; n=10/40, 25%)
- Therapeutic alliance (Enlight, APA; n=3/40, 8%)

Health outcomes

- Health benefits and effectiveness (ISAT, RE-AIM, TEACH-apps, RACE, A-MARS, EUEnHTA, EVALAPPS; n=15/40, 38%)
- Patient safety (Pre-Cert, LCDH, IOM, MARS, NICE ESF, EUEnHTA; n=15/40, 38%)
- Evidence base (Enlight, APA, A-MARS, NICE ESF, RACE, LCDH; n=11/40, 28%)

Visible popularity metrics

- Ratings and user satisfaction (TEACH-apps, MARS, NICE ESF; n=12/40, 30%)

Social aspects

- Social influence and endorsement (EUEnHTA, LCDH; n=7/40, 18%)

Sustainability and scalability

- Cost-effectiveness (ISAT, RE-AIM, APA, BIT, IOM, NICE ESF, EUEnHTA; n=17/40, 43%)
- Maintenance (ISAT, RE-AIM, REP, TEACH-apps, BIT, CLIQ; n=13/40, 33%)
- Accessibility and fidelity (ISAT, RE-AIM, BIT, NICE ESF, EUEnHTA, RACE; n=8/40, 20%)
- Availability (EVALAPPS; n=2/40, 5%)

Health care organization

- Implementation (RE-AIM, TEACH-apps; n=4/40, 10%)
- Workforce and resources (ISAT, EUEnHTA; n=3/40, 8%)
- Infrastructure (ISAT, EUEnHTA; n=3/40, 8%)

Health care context

- Strategic, political, and environmental contexts (ISAT, TEACH-apps, REP; n=3/40, 8%)

Developer

- Transparency and credibility (APA, Pre-Cert, MedAd-AppQ; n=11/40, 28%)
- Compliance and accountability (Pre-Cert, EUEnHTA; n=7/40, 18%)
- Proactivity and interaction quality (Pre-Cert; n=2/40, 5%)
- History of producing safe health products (Pre-Cert; n=1/40, 3%)

and relevance (MedAd-AppQ, Enlight, A-MARS, NICE ESF, and EVALAPPS; 15/40, 38%), meaning appropriate and relevant features to meet the clinical aim, the right mix of ability and motivation, and meeting the intended purpose; and feature convenience (MedAd-AppQ; 3/40, 8%), which typically assesses how convenient or bothersome some of the features are, such as reminders, push notifications, and daily prompts.

The content cluster includes content quality (Enlight, A-MARS, CLIQ, MARS, A-MARS, and EVALAPPS; 17/40, 42%), which assesses the quality of the health-related content (accuracy, completeness, consistency, and timeliness); content credibility (Enlight, APA, CLIQ, MARS, A-MARS, and NICE ESF; 15/40, 38%), which looks into content source credibility (eg, the WHO), advisory support, third-party verification, or the level of clinicians’ involvement in the tool’s content development; and content validity and reliability (MedAd-AppQ, NICE ESF, and EVALAPPS; 10/40, 25%), typically defined as the extent to which a tool’s contents are relevant to the underlying construct and likely to be effective in achieving a particular intervention purpose in a specific intended population.

The data management cluster includes data privacy and security (FDA precertification program [Pre-Cert], MedAd-AppQ, Enlight, APA, CLIQ, LCDH, NICE BCG, MARS, NICE ESF, RACE, EUEnHTA, and EVALAPPS; 26/40, 65%)—which assess the cybersecurity responsibility, presence of healers, informed consent, and privacy policy and whether the treatment
of any data is compatible with the Patient Data Act, Personal Data Act, and other applicable privacy laws—and data integration and interoperability (APA and A-MARS; 7/40, 18%), which evaluate the tool’s ability to exchange information with and use information from other health technologies (eg, electronic health records) and users’ ability to smoothly move across different platforms.

The design cluster includes the tool’s usability (Enlight, APA, BIT, AQEL, CLIQ, IOM, NICE BCG, RACE, EUENetHTA, and EVALAPPS; 27/40, 68%), which assesses user experience, navigation, learnability, and ease of use; visual design (Enlight, MARS, and A-MARS; 12/40, 30%), which evaluates esthetics, layout, size, pop-up windows and flash images, visual appeal, and consistency of the theme throughout the tool; and timeliness (IOM and A-MARS; 4/40, 10%), typically defined as the ability to use the tool in real time (ie, real-time data tracking), reducing waits and sometimes harmful delays for both those who receive and those who provide care.

Social Assessment Criteria

The social assessment criteria were grouped into 4 clusters: human centricity, health outcomes, visible popularity metrics, and social aspects. The human centricity cluster includes user engagement, customizability, tailoring, and user control (Enlight, REP, TEACH-apps, NICE BCG, MARS, A-MARS, and RACE; 17/40, 42%), meaning the tool’s interactivity and the ability to enable customization, collaboration, participation, information sharing, and decision-making in one’s own health as well as evidence for collaboration with users; behavior change and persuasiveness (Enlight, AQEL, NICE BCG, NICE ESF, and RACE; 14/40, 35%), which assess whether the tool reflects a persuasive design that aims to understand what influences people’s behavior and decision-making and then uses this information to design compelling user interactions (call for action, load reduction of activities, therapeutic rationale and pathway, rewards, real data-driven and adaptive, and ongoing feedback); equity, accessibility, and inclusiveness (IOM; MARS, A-MARS, NICE ESF, and EUENetHTA; 10/40, 25%), which look into whether the tool supports providing care that takes the user context into account and does not vary in quality because of personal characteristics such as gender, ethnicity, geographic location, and socioeconomic status (eg, tools that are accessible to vulnerable populations such as people with disabilities, patients with chronic diseases, patients with mental illnesses, pediatric patients, maternity patients, and older adults); and therapeutic alliance (Enlight and APA; 3/40, 8%), defined as the tool’s ability to foster interaction between clinicians and their patients.

The health outcomes cluster includes health benefits and effectiveness (ISAT, RE-AIM, TEACH-apps, IOM, NICE BCG, NICE ESF, and EUENetHTA; 15/40, 38%), which typically assess health benefits of the new technology in producing health benefits in a real-world setting, also referred to as real-world evidence; patient safety (Pre-Cert, LCDH, IOM, MARS, NICE ESF, and EUENetHTA; 15/40, 38%), which looks into the ability of an eHealth tool to handle “dangerous” information entered by a patient and avoid safety risks to patients from the care that is intended to help them; and evidence base (Enlight, APA, A-MARS, NICE ESF, RACE, and LCDH; 11/40, 28%), which reflects the presence of solid scientific evidence supporting the tool’s health claims (eg, published research and randomized controlled trials).

The visible popularity metrics cluster includes ratings and user satisfaction (TEACH-apps, MARS, and NICE ESF; 12/40, 30%), which reflect users’ perceived value through users’ reviews and ratings (as a proxy for quality, usefulness, or acceptability and popularity). Finally, the social aspects cluster includes social influence and endorsement (EUENetHTA and LCDH; 7/40, 18%), which assess the possibilities for peer support, social networking, information sharing, and endorsement by health care professionals.

Organizational Assessment Criteria

The organizational assessment criteria were grouped into 4 clusters: sustainability and scalability, health care organization, health care context, and developer. The sustainability and scalability cluster includes cost-effectiveness (ISAT, RE-AIM, APA, BIT, IOM, NICE ESF, and EUENetHTA; 17/40, 42%), which assesses the balance between the costs and benefits arising from the tool’s use. This refers to the tool’s direct costs (eg, purchase price, subscription, and licensing) but may also include costs associated with the tool’s selection, staff training, setting up support mechanisms, and appropriate governance. This cluster also includes maintenance (ISAT, RE-AIM, REP, TEACH-apps, BIT, and CLIQ; 13/40, 32%), which assesses the commitment of the developers to maintaining their products in the long term by conducting periodic updates and maintenance (from both technical and content perspectives); adoption and fidelity (ISAT, RE-AIM, BIT, NICE ESF, EUENetHTA, and RACE; 8/40, 20%), which look into the tool’s adoption rates, acceptability, and desirability as well as its integration into clinical practice, system use, and adherence; and availability (EVALAPPS; 2/40, 5%), which evaluates the guarantee of access to the tool and its data at any time and its availability on different operating systems (eg, Android and iOS).

The health care organization cluster includes implementation (RE-AIM and TEACH-apps; 4/40, 10%), which assesses the extent to which the intervention was delivered as intended (eg, feasibility of delivering all components of an intervention at a predetermined date and time); workforce and resources (ISAT and EUENetHTA; 3/40, 8%), which assess the workforce required to scale up the tool and the implications for care processes and care management; and infrastructure (ISAT and EUENetHTA; 3/40, 8%), which assesses the readiness of the necessary infrastructure for the tool’s implementation. The health care context cluster includes strategic, political, and environmental contexts (ISAT, TEACH-apps, and REP; 3/40, 8%) and evaluates how favorable are the preconditions (strategic, political, and environmental contexts) that influence the scaling up of the eHealth tool, for example, the intervention’s suitability to the socioeconomic context in question, considerations of foreign languages that the tool needs to support, literacy level, and the local regulatory environment.

The developer cluster includes the transparency and credibility of the tool’s developer (APA, Pre-Cert, and MedAd-AppQ; 11/40, 28%), which look into the availability of information.
and credentials of the individuals and organizations involved in the development and funding of the tool; compliance and accountability (Pre-Cert and EUNetHTA; 7/40, 18%), which assess the developer’s ethical conduct, clinical responsibility, and respect for the rules and regulations protecting patients’ rights and societal interests; proactivity and interaction quality (Pre-Cert; 2/40, 5%), which evaluate the interaction quality between the provider and the users, including responsiveness, after-sales services, and customer orientation as well as the demonstration of excellence in a proactive approach to the assessment of user needs and continuous learning; and, finally, the history of producing safe health products (Pre-Cert; 1/40, 2%), which assesses whether the developer has successfully delivered safe health products in the past.

**Discussion**

**A Scattered and Fragmented Landscape**

Although there are various initiatives working on finding ways to assess the quality of eHealth tools, these efforts face multiple challenges, as shown in the overview in **Figure 4**.

**Comparability**

The multitude of frameworks and initiatives attempting to address the topic of eHealth tool assessment shows the lack of standardization in this field and adds another challenge for the relevant stakeholders as they are faced with proliferating approaches and not knowing which assessment tool to use or how best to use it [98,99]. The diversity of assessment methods sometimes results in a lack of clarity or comparability [20,30,32,35,36,39]; furthermore, this scattered landscape also signals the lack of generalizability and standardization in this field of research [32]. Moreover, assessment and data collection methods vary widely between the different initiatives (eg, self-reported vs objective measures and qualitative vs quantitative assessment) [34,37,39,50,98-100].

**Practicability**

In many cases, there is limited information and methods describing how to realistically assess and evaluate these tools in practice [19,33]; many of the existing initiatives are conceptual without granular guidance on how to use and apply them in day-to-day decision-making [37,56,92]. For instance, the work by Kloc et al [101] compared the English NICE ESF for digital health technologies and the French National Authority for Health guide on the assessment of connected medical device guidelines and concluded that the guidelines do not always clearly describe the assessment process or the specific criteria determining the decision. Correspondingly, Bradway et al [99] suggested that users should be provided with guidance and educational resources on how to perform a proper assessment.

**Criteria Completeness**

Moreover, research has shown that some of the existing initiatives sometimes overlook important assessment criteria, resulting in incomplete or issue-specific assessment formworks [32,35,36,51,99].

**Regulatory Complexity**

The lack of regulatory clarity and the absence of institutionalized quality controls in many countries make a comprehensive definition of the assessment criteria more challenging [15,41-43,53]. Moreover, there are some shortcomings with some of the current certification labels, as highlighted by Bradway et al [99], who pointed out that, even though common labels may categorize a tool as a medical device, it may still include the warning in fine print that it is intended for...
entertainment only, showing a lack of accountability and creating confusion on the users’ side. There are also many gray areas in existing regulatory oversight efforts; for instance, the US FDA applies regulatory oversight only to a small subset of tools that qualify as medical devices and potentially pose a risk to patient safety [9,102]. The European regulatory system offers another model in which each member state can file an approval application for a high-risk medical device and obtain a Conformité Européenne mark. However, although Conformité Européenne marks indicate that these tools are compliant with European legislation, the tools only need to demonstrate safety and performance but not clinical efficacy [102]. These regulatory gaps mean that the safety, efficacy, and ethical compliance of certified eHealth tools cannot be guaranteed, posing a potential threat to patients’ safety [103].

Validation
Furthermore, the validity and reliability of the existing assessment tools and frameworks have not always been rigorously tested [17,50,56]; such validation efforts are key to ensure assessment processes that reflect the real-world needs of the different stakeholders in the health care ecosystem [17].

Contextuality
Relatively, eHealth interventions are highly contextual, making it crucial to consider the implementation context and use cases, but the varying contexts and use cases make it quite challenging to find a standardized and generalizable way to assess them [15,17,18,100].

Information Availability
Proper assessment is mostly dependent on developer transparency and the availability of information, which is unfortunately not always the case, making it quite challenging to address the questions needed to accurately assess the quality and impact of an eHealth tool [9,98]. Concerningly, a previous study showed that, in a sample of 52 eHealth tools, 63.5% of the providers gave no information about the tool itself, 67.3% did not provide information about the credentials of the developers or consultants, and only 4% provided information supporting the tool’s efficacy [104].

Subjective Measures
Although most assessment criteria are objective, some of the criteria that are most relevant for user engagement are subjective, as pointed out by Lagan et al [98], limiting the standardization of the assessment outcome. For example, given the importance of user engagement for the success of eHealth tools [23,24,105,106], it would still be crucial to include assessment criteria that reflect key user engagement and adoption drivers such as ease of use and visual appeal [23,24,63,107].

Assessor Diversity
In addition, as Bradway et al [99] noted, some assessment initiatives do not involve or even inform all the relevant stakeholders of assessment results, establishing the importance of involving diverse assessor profiles, including the tools’ developers themselves.

**Tool’s Life Cycle**
Finally, most existing assessment frameworks focus only on eHealth tools that are fully operational within the market and do not necessarily tackle those that are still under development or have not been implemented yet [99]. One of the few assessment frameworks that look into specific criteria for the different phases of the development and implementation cycle is the framework for the design and evaluation of digital health interventions developed by Kowatsch et al [108] categorizing the assessment criteria according to the phase in which the tool is in terms of preparation, optimization, evaluation, and implementation.

It is worth noting that most national initiatives are also still in their infancy and facing several teething problems, which shows that these frameworks have not reached a high enough maturity level yet. For instance, even though Germany became the first country worldwide to approve certain eHealth tools, referred to as Digitale Gesundheitsanwendungen (DiGA) in German, meaning digital health applications, for prescription with costs covered by standard statutory health insurance, research has shown that clinicians’ adoption rates of this option are still rather low [109]. Similarly, the FDA has recently announced that its Pre-Cert program, which focuses on medical technology providers and their internal processes rather than on individual devices and apps, is still not ready to go beyond the pilot phase [110,111]. In addition, Alon et al [15] stated that they were unable to identify a standard measure that differentiated the tools requiring regulatory review from those that did not when they assessed the Pre-Cert program.

Despite these challenges, efforts to harmonize and standardize assessment approaches are ongoing. For example, the European Committee for Standardization (CEN) International Organization for Standardization (ISO) technical specification for the quality and reliability of health and wellness apps (CEN ISO/TS 82304-2), published in 2021, provides quality requirements for health apps and defines a health app quality label to visualize the quality and reliability of these apps [112]. Horizon Europe project “Label2Enable” involves 14 organizations from 7 countries (Belgium, Croatia, Germany, Italy, Lithuania, the Netherlands, and Spain) that have joined forces to promote the CEN ISO/TS 82304-2 health app assessment framework and label in Europe [113].

**The Relevance of a Sociotechnical Approach to eHealth Assessment**
Despite the multitude of initiatives attempting to address this topic, it remains that there are multiple challenges to be addressed. It is also clear that developing a comprehensive assessment criteria framework for eHealth will be challenging owing to its multidimensional nature [19,41-43]. The findings from this systematic review show that there is no single framework that is used uniformly to evaluate the different assessment criteria of eHealth tools. However, it is worth noting that, despite their different contexts and the different disease conditions they addressed, there was substantial overlap among the frameworks. Nevertheless, although these initiatives attempt to provide relevant information on the quality of eHealth tools, they are not always able to address all stakeholder issues, and
although most criteria can be related to one framework or another, no framework seems to cover all relevant criteria without being extended.

We propose an aggregated framework adopting a sociotechnical approach to eHealth evaluation balancing the technical, social, and organizational assessment criteria. This aggregated framework considers all the criteria appearing in the included studies and classifies them according to the sociotechnical framework; this aggregation should help overcome some of the identified challenges with current efforts, namely, incomplete assessment measures [114]. Our approach also acknowledges that health care technology cannot be successfully implemented and scaled in isolation from the broader organizational and social contexts in which it is being used and that, therefore, we need to use frameworks that consider implementation challenges in light of the complexity of the sociotechnical structure and interplay between the technical, social, and organizational aspects. Figure 5 summarizes our proposed aggregated framework that considers all the criteria covered in the included studies, classifying them according to the sociotechnical framework. The arrows in the figure indicate the continuity and interconnectedness between the social, organizational, and technical criteria.

**Figure 5.** Sociotechnical framework to assess the quality and impact of eHealth tools.

Technical assessment criteria are the foundation for the viability of any eHealth solution and for it to be considered at all by potential users; without this foundation, a tool would not properly meet the basic requirements for success. This is most likely why technical aspects have mostly been the focus of existing initiatives and frameworks [51]. For instance, the only assessment criteria that were reflected in more than half of the included studies were the tools’ usability (27/40, 68%) and data privacy and security (26/40, 65%), highlighting the current focus on assessing the technical aspects without necessarily...
giving enough weight to social and organizational assessment measures, as demonstrated in our previous discussion. This was similarly highlighted by Lagan et al [98], who pointed out the rising popularity of data privacy criteria in assessment frameworks in recent years.

Ensuring a high level of technical performance and offering well-defined and useful functionalities and features as well as credible, valid, and reliable content; proper data management strategies; and a superior user experience are the basics that every eHealth tool must meet for it to be considered by the relevant users. Even though feature usefulness may seem like an intuitive and basic requirement for the success of any eHealth technology, Singh et al [54] reported that their evaluation of 143 tools targeting patients who have high needs and incur high health care costs showed that only a minority of these tools appeared likely to be useful to patients.

It is also worth noting that, although data integration and interoperability were only mentioned in 18% (7/40) of the included studies, previous studies have shown that this is an important user requirement. User adoption research has shown that interoperability issues can raise clear concerns when eHealth tools cannot be integrated into the hospital’s or clinic’s current systems or when there are limitations in data integration and exchange [23,63]. This technical criterion closely affects and is affected by the organizational criteria related to infrastructure and implementation. It is also closely related to the sustainability and scalability organizational criteria, showing the interconnectedness between these elements that contribute to the potential success of a given eHealth tool.

The inclusion of organizational assessment criteria may help address a key challenge with current efforts related to the importance of the contextuality of eHealth tools as these technologies are not used in isolation of the health care ecosystem; therefore, a proper assessment of the potential impact of these tools should consider the specific context. Health care technologies are generally more complex than tools that address individual user needs as they usually support patients with comorbidities who are typically treated by multidisciplinary teams that might even work in different health care organizations, hence the importance of contextual and organizational aspects to assess the potential impact of these novel solutions. Context-specific criteria such as implementation, workforce and resources, infrastructure, and the overall health care context do not seem to be fairly represented in the current assessment initiatives. Our analysis showed that only 10% (4/40) of the included studies encompassed implementation criteria, and only 8% (3/40) looked into the required infrastructure, workforce, and resources as well as social, political, and environmental contexts. This results in situations where a tool may be of good quality when assessed in isolation but might not have the desired impact in a real-life scenario because of contextual criteria that do not necessarily allow it to be successfully implemented or scaled if not properly evaluated.

To put things into perspective, it is important to consider the factors affecting user adoption when assessing potential eHealth tools to avoid situations where a tool may be of good quality in isolation of its context but not a good fit when rolled out in a real-life setting. A comprehensive systematic review that looked into the factors affecting clinician adoption of eHealth tools in 171 published studies indicated that organizational factors, especially workflow-related factors such as implications for the workload and workflow, the infrastructure required for the implementation, and the wider health care context such as local regulations, are crucial for clinician adoption [23,63], showing some disconnect between the focus of the current assessment efforts and what it takes for a tool to be successfully adopted by its intended users in a real-life context.

Even though the availability of information is one of the challenges facing current initiatives, as explained in the previous section, less than one-third (11/40, 28%) of the included studies incorporated organizational assessment criteria regarding the developers’ transparency and credibility. Our approach proposes the inclusion of developer-related criteria by evaluating the developers’ transparency and credibility, compliance and accountability, proactivity and interaction quality, and history of producing safe tools to help overcome this challenge and entice tool providers to transparently communicate the information needed for their very own assessment.

Hence, the overall organizational assessment criteria should comprise criteria regarding the sustainability and scalability of the tool (cost-effectiveness, maintenance, adoption and fidelity, and availability); criteria related to health care organizations in the specific context being assessed (implementation, workforce and resources, and infrastructure); criteria related to the wider health care context, such as local regulations and certification requirements; and criteria to assess the developers’ credibility, compliance, and interaction quality.

We equally advocate for the importance of the inclusion of relevant social assessment criteria that evaluate the potential societal impact of these tools. Notably, even though many frameworks included usability in general as an assessment criterion, more than half (23/40, 58%) of the included studies did not specifically address human centrality through active user engagement and behavior change strategies. This is concerning considering the lack of reliable evidence regarding the ability of most commercially available eHealth tools to induce lasting behavior change [99,115]. Proper user engagement and effective behavior change design strategies may help address issues reported in previous studies that established that only a small fraction of patients kept using eHealth tools in the long term and that up to 80% of users would only show minimal engagement, using the tools <2 times [116,117]. Another study conducted on a large real-world cohort of 189,770 people reported that only 2.6% of the people who downloaded an eHealth tool sustained its active use [118], concluding that the impact of such tools may remain minimal if they fail to properly engage patients, making this a vital assessment criterion. Although developers seem to pay less attention to behavior maintenance than to initiation and evidence for collaboration with users or professionals is mostly lacking, as reported by McMillan et al [48], promisingly, Baumel et al [31] noted some advancements made in recent years as human-centric criteria related to persuasive design and therapeutic alliance gain more importance. This social criterion
closely affects and is affected by the technical criteria related to a tool’s design and usability.

Nonetheless, 75% (30/40) of the included studies failed to address some core social principles, such as the equity, accessibility, and inclusiveness of the tools being assessed, overlooking the vital societal impact of such criteria. We highlight the importance of the inclusion of these measures as inclusive design principles may help developers address the needs of the most susceptible patient populations who may not be engaging with such technologies owing to their age, health-related physical and cognitive challenges, educational level, socioeconomic status, or technological skills and experience [24]. Designing for inclusivity does not ignore the unique features, environments, and cultural contexts of users; many aspects of the digital divide may be addressed through an inclusive design that incorporates cultural appropriateness, easy-to-understand lay language that does not require high literacy levels, and ease of use that does not require any sophisticated technical skills [24]. Unfortunately, equity seems to be one of the less frequently observed criteria in eHealth tools, as equally reported by Lee et al [46] and confirmed by our findings. Assessing such criteria would increase the chances of having tools that are designed in a way that makes them more accessible to the very patients who need them the most.

Surprisingly, less than 40% (15/40, 38%) of the included studies considered criteria related to health outcomes, such as health benefits and effectiveness, patient safety, and evidence base. This may affect the societal impact of these tools if not assessed when determining a specific tool’s potential impact on health, which is supposed to be the main reason why people use these tools, especially when previous studies have indicated that the clinical benefit of many of these tools is quite limited or insufficient, as reported by Huckvale et al [91]. This social criterion is closely affected by the technical criteria related to a tool’s features and content.

Generally, comprehensive social assessment criteria according to our findings should encompass human centricity (by assessing user engagement, customizability, behavior change strategies, the tool’s inclusiveness, and its impact on the therapeutic alliance), health outcomes (by assessing health benefits and effectiveness, patient safety, and evidence base), visible popularity metrics such as tool ratings and user satisfaction, and other influential aspects such as social influence and endorsement.

Limitations and Future Research

This study contributes to the understanding of the different criteria used to assess the quality and impact of eHealth tools; however, some limitations must be acknowledged. This review may not have included relevant studies that were not indexed in the searched databases or were written in a language other than English as well as gray literature searches that could have also allowed for the identification of additional relevant insights. However, this study focused on peer-reviewed scientific papers. In addition, this analysis only considered published studies, and no further contact was made with the authors of the papers to obtain additional information or validate our thematic analysis. We also did not include articles based on manual searches of reference lists to avoid a biased sample of studies given that positive studies are more likely to be cited. Consequently, it is possible that other frameworks, initiatives, or assessment criteria were missed.

Future work could include studies in other languages to gain a better grasp of any interregional or Intercultural differences. The authors also intend to build on this review by conducting another study to critically apply, reflect, validate, and revise the criteria aggregated in this study with the relevant stakeholders and co-create accessible and easy-to-use tools with practice experts that may support them in their eHealth assessment decisions.

Conclusions

The findings from this systematic review demonstrate that there is no single framework that is used uniformly to assess the quality and impact of eHealth tools. Current assessment efforts face some core challenges, such as the lack of comparability and practicability, gaps in criteria completeness of the individual frameworks, regulatory complexity, issues with the validation of existing frameworks, the contextuality of eHealth tools, the availability of the information necessary for the assessment, the need to include subjective measures, and the lack of assessor diversity in many cases. This review also highlights the need for a more comprehensive approach that balances the social, organizational, and technical assessment criteria in a way that reflects the complexity and interdependence of the health care ecosystem and is aligned with the factors affecting users’ adoption to ensure uptake and adherence in the long term.

Our proposed framework aggregates and expands the criteria appearing in the included studies and classifies them according to the sociotechnical framework, acknowledging that health care technologies cannot be successfully implemented and scaled in isolation from the broader organizational and social contexts in which they are being used and that, therefore, we need to use frameworks that consider implementation challenges in light of the complexity of the sociotechnical structure and interplay between the technical, social, and organizational aspects. More efforts are needed to find ways to overcome the identified challenges and validate the aggregated framework resulting from this study with the relevant stakeholders to ensure its pertinence and help make it more usable and accessible to potential assessors to support a more comprehensive process of evaluating the quality and impact of eHealth technologies.

Acknowledgments

This study is part of an overarching research project that is jointly sponsored by F. Hoffmann-La Roche Ltd, KPT insurance, and Innosuisse (the Swiss Innovation Agency; grant 104.445 IP-ICT).
Conflicts of Interest

CJ is an editorial board member of JMIR Human Factors at the time of this publication. The other authors have no conflicts of interest to declare.

Multimedia Appendix 1
Critical Appraisal Skills Programme checklist.

[PDF File (Adobe PDF File), 440 KB - humanfactors_v10i1e45143_app1.pdf]

Multimedia Appendix 2
Critical appraisal of the included studies.

[XLSX File (Microsoft Excel File), 66 KB - humanfactors_v10i1e45143_app2.xlsx]

Multimedia Appendix 3
Phases of thematic analysis based on the work by Braun and Clarke [31-33].

[PDF File (Adobe PDF File), 170 KB - humanfactors_v10i1e45143_app3.pdf]

Multimedia Appendix 4
Characteristics of the included studies.

[PDF File (Adobe PDF File), 101 KB - humanfactors_v10i1e45143_app4.pdf]

Multimedia Appendix 5
Frameworks and guidelines that resulted from or were used in the included studies.

[PDF File (Adobe PDF File), 134 KB - humanfactors_v10i1e45143_app5.pdf]

Multimedia Appendix 6
Assessment criteria and their occurrence, with references.

[PDF File (Adobe PDF File), 170 KB - humanfactors_v10i1e45143_app6.pdf]

References


9. Larson RS. A path to better-quality mHealth apps. JMIR Mhealth Uhealth 2018 Jul 30;6(7):e10414 [FREE Full text] [doi: 10.2196/10414] [Medline: 30061091]


63. Jacob C, Sanchez-Vazquez A, Ivory C. Understanding clinicians’ adoption of mobile health tools: a qualitative review of the most used frameworks. JMIR Mhealth Uhealth 2020 Jul 06;8(7):e18072 [FREE Full text] [doi: 10.2196/18072] [Medline: 32442132]


73. Jacob C, Sanchez-Vazquez A, Ivory C. Understanding clinicians’ adoption of mobile health tools: a qualitative review of the most used frameworks. JMIR Mhealth Uhealth 2020 Jul 06;8(7):e18072 [FREE Full text] [doi: 10.2196/18072] [Medline: 32442132]


85. UsabilityNet. 2006. URL: https://usabilitynet.org/ [accessed 2022-03-01]


Abbreviations

A-MARS: adapted Mobile App Rating Scale
APA: American Psychiatric Association
AQEL: App Quality Evaluation
BIT: Behavior Interventions Using Technology
CASP: Critical Appraisal Skills Programme
CEN: European Committee for Standardization
CLIQ: Clinical Information Quality
DiGAt: Digitale Gesundheitsanwendungen
EUNetHTA: European Network for Health Technology Assessment
FDA: Food and Drug Administration
IOM: Institute of Medicine
ISAT: Intervention Scalability Assessment Tool
ISO: International Organization for Standardization
LCDH: Legal Challenges in Digital Health
MARS: Mobile App Rating Scale
MedAd-AppQ: Medication Adherence App Quality
NICE BCG: National Institute for Health and Care Excellence behavior change guidance
NICE ESF: National Institute for Health and Care Excellence Evidence Standards Framework
PENG: Prioritering efter NyttoGrunder
Pre-Cert: Food and Drug Administration precertification program
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RACE: Review, Assess, Classify, and Evaluate
RE-AIM: reach, effectiveness, adoption, implementation, and maintenance
REP: Replicating Effective Programs
TEACH-apps: Technology Evaluation and Assessment Criteria for Health Apps
WHO: World Health Organization

©Christine Jacob, Johan Lindeque, Alexander Klein, Chris Ivory, Sabina Heuss, Marc K Peter. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 23.03.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Review

Feasibility and Acceptability of Chatbots for Nutrition and Physical Activity Health Promotion Among Adolescents: Systematic Scoping Review With Adolescent Consultation

Rui Han¹, BSc, MPH; Allyson Todd¹, BGS, MPH; Sara Wardak¹; Stephanie R Partridge¹, BSc (Hons I), APD, PhD; Rebecca Raeside¹, BBiomedSc, MPH

Engagement and Co-Design Research Hub, School of Health Sciences, Faculty of Medicine and Health, University of Sydney, Westmead, Australia

Abstract

Background: Reducing lifestyle risk behaviors among adolescents depends on access to age-appropriate health promotion information. Chatbots—computer programs designed to simulate conversations with human users—have the potential to deliver health information to adolescents to improve their lifestyle behaviors and support behavior change, but research on the feasibility and acceptability of chatbots in the adolescent population is unknown.

Objective: This systematic scoping review aims to evaluate the feasibility and acceptability of chatbots in nutrition and physical activity interventions among adolescents. A secondary aim is to consult adolescents to identify features of chatbots that are acceptable and feasible.

Methods: We searched 6 electronic databases from March to April 2022 (MEDLINE, Embase, Joanna Briggs Institute, the Cumulative Index to Nursing and Allied Health, the Association for Computing Machinery library, and the IT database Institute of Electrical and Electronics Engineers). Peer-reviewed studies were included that were conducted in the adolescent population (10-19 years old) without any chronic disease, except obesity or type 2 diabetes, and assessed chatbots used nutrition or physical activity interventions or both that encouraged individuals to meet dietary or physical activity guidelines and support positive behavior change. Studies were screened by 2 independent reviewers, with any queries resolved by a third reviewer. Data were extracted into tables and collated in a narrative summary. Gray literature searches were also undertaken. Results of the scoping review were presented to a diverse youth advisory group (N=16, 13-18 years old) to gain insights into this topic beyond what is published in the literature.

Results: The search identified 5558 papers, with 5 (0.1%) studies describing 5 chatbots meeting the inclusion criteria. The 5 chatbots were supported by mobile apps using a combination of the following features: personalized feedback, conversational agents, gamification, and monitoring of behavior change. Of the 5 studies, 2 (40.0%) studies focused on nutrition, 2 (40.0%) studies focused on physical activity, and 1 (20.0%) focused on both nutrition and physical activity. Feasibility and acceptability varied across the 5 studies, with usage rates above 50% in 3 (60.0%) studies. In addition, 3 (60.0%) studies reported health-related outcomes, with only 1 (20.0%) study showing promising effects of the intervention. Adolescents presented novel concerns around the use of chatbots in nutrition and physical activity interventions, including ethical concerns and the use of false or misleading information.

Conclusions: Limited research is available on chatbots in adolescent nutrition and physical activity interventions, finding insufficient evidence on the acceptability and feasibility of chatbots in the adolescent population. Similarly, adolescent consultation identified issues in the design features that have not been mentioned in the published literature. Therefore, chatbot codesign with adolescents may help ensure that such technology is feasible and acceptable to an adolescent population.
Introduction

Adolescents, aged 10-19 years, as defined by the World Health Organization (WHO), are a unique age group, who begin to develop independent lifestyle habits that they carry into adulthood [1]. Concerning the prevalence of overweight and obesity among adolescents is increasing worldwide. In 2016, more than 31 million children and adolescents aged 5-19 years were reported as overweight or obese [1]. Overweight and obesity in adolescence are associated with poorer health outcomes in adulthood, including cardiovascular disease and type 2 diabetes [2]. Therefore, intervening early in the life course is critical to prevent the future burden of chronic disease and comorbidities [3,4]. Regular physical activity and optimal nutrition are fundamental in preventing and assisting those with overweight and obesity to return to a healthy weight. Worldwide, more than 80% of adolescents do not meet the recommended levels of physical activity or sedentary behavior guidelines [4]. Since the COVID-19 pandemic began, research has reported increased screen time being associated with weight gain among adolescents [2]. Additionally, most adolescents fail to meet WHO’s guidelines on daily fruit and vegetable intake [5]. The overconsumption of nutrient-poor, ultraprocessed foods and sugar-sweetened beverages is further contributing to the rising rates of overweight and obesity. Simultaneously, malnutrition, micronutrient deficiencies, and food insecurity continue to persist among adolescents worldwide [6]. Adolescents need support to improve physical activity and nutrition behaviors, which in turn will minimize the growing rate of adolescents with overweight and obesity worldwide.

Digital health interventions, such as mobile apps, text messaging, and gamification, show promise for improving the health of adolescents through targeting physical activity and dietary behaviors [7,8]. Nearly 70% of adolescents in high-income countries have a smartphone and are frequently online [9]. Mobile-based interventions are relatively low cost, accessible, and widely acceptable among adolescents [10]. Gamification is the implementation of game design elements in real-world contexts for nongaming purposes [11] and has been found to be effective in improving physical activity levels, fruit and vegetable intake, and nutrition knowledge in adolescents [12,13]. For example, the popular online game Pokémon Go has been found to promote physical activity [7,14]. Mobile apps may assist in improving adolescents’ health with a plethora of apps available. A review by Schoeppe et al [15] found that currently available mobile apps that promote physical activity and nutrition have moderate quality and use a range of behavior change techniques, such as encouragement, performance feedback, and gamification. However, there is limited knowledge of user engagement [15]. A randomized controlled trial, conducted in 14 secondary schools in Australia, evaluated the influence of a mobile app to promote physical activity in adolescents and found that half of the participants were influenced by the “push-prompt” message reminder to be active, reduce sweetened beverage consumption, and reduce screen time [16]. Further, the use of semipersonalized text messaging has been found to be a feasible and acceptable strategy to engage adolescents to promote healthy behaviors [17]. Incorporating gamification and personalized feedback may help improve engagement for young people in digital health interventions [13]. As technology continues to evolve, it is important to evaluate emerging features to help improve and sustain diet and physical activity behaviors among adolescents [7,18].

Artificial intelligence (AI) is a rapidly developing technical science being applied to the healthcare field [8,19]. It is commonly used in precision medicine, using machine learning, which involves training models with data [19]. The use of natural language processing (NLP) allows AI to communicate using human-like language, as well as to extract and construct information from social media and medical documents [8]. AI items, such as Apple Siri and Google Assistant, are becoming increasingly popular among the public to answer health-related questions [20,21]. Chatbots are an emerging software application designed for text-based conversation. They can search for information from the internet or a database to respond to users’ inquiries and personalize communication with humans [22]. Chatbots can be designed with or without AI. Those without AI cannot learn and adapt and often have predetermined responses based on the question asked by the user. However, AI chatbots are trained to have human-like conversations using NLP. Therefore, there is potential for the use of chatbots as a digital health intervention to improve nutrition and physical activity behaviors across the life course. There is current evidence of chatbots promoting physical activity in the adult population, which is encouraging, but further research is needed to support these findings [23]. A systematic review investigating the use of chatbots to improve physical activity and nutrition across all age groups found no studies specifically targeting adolescents [23]. Chew’s [24] recent scoping review of chatbots used to promote weight loss across all age groups also found the same gap in knowledge and highlighted the importance of using age-appropriate design features to enhance engagement for adolescents. There is potential for this cost-effective and highly accessible technology to deliver health information to young people to improve their nutrition and physical activity behaviors [25]. However, there is limited research on the feasibility and acceptability of chatbots in the adolescent population [23]. This systematic scoping review aims to evaluate findings from peer-reviewed, published studies to understand the feasibility and acceptability of chatbots to promote nutrition and physical activity in adolescents. A secondary aim is to identify design features of chatbots that would be acceptable and feasible with an established youth advisory group.
Methods

Study Design
A scoping review was determined to be the most suitable method to synthesize data to identify knowledge gaps and look broadly at the existing literature [26]. The systematic scoping review methodology was informed by the 6-stage methodological framework outlined by Arksey and O’Malley [27] and the Joanna Briggs Institute guidelines for scoping reviews [28]. The review was reported following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for scoping reviews (PRISMA-ScR) [29]. The research questions were formulated by the research team, along with the eligibility criteria for including relevant studies. Next, studies were selected based on the predefined eligibility criteria, and relevant data from the included studies were extracted. Following data extraction, results were collated and summarized narratively.

Eligibility Criteria
To be included, peer-reviewed research studies must have (1) been conducted in the adolescent population, defined according to WHO as the second decade of life (10-19 years); (2) participants without a chronic disease, except obesity or type 2 diabetes; (3) assessed the feasibility and acceptability of chatbots used for nutrition or physical activity interventions or both that encourage individuals to meet dietary or physical activity guidelines and support positive behavior change; (4) been conducted in 2010 and beyond (to coincide with the period that smart devices were normalized in society, including chatbots); and (5) been written in any language and conducted in any country. Quantitative and qualitative peer-reviewed papers were included. For this study, chatbots were defined as programs that contained a conversational agent that could engage in “small talk”; smart conversational agents, such as Apple Siri; and those involving a computer-generated virtual agent.

Search Strategy
Initially, a limited search of Google and MEDLINE was completed by the authors to evaluate the scope of existing research in the literature. The search strategy was developed in conjunction with the academic liaison librarian. An advanced search was conducted in March 2022 using MEDLINE, including Medical Subject Headings (MeSH) and keyword searches, in 3 core concept areas: chatbots, nutrition intervention, and physical activity intervention. An extensive list of synonyms for all terms was included to capture the maximum number of studies (Multimedia Appendix 1). Once key concepts and terms were determined, the search strategy was adapted to other database searches. The search was implemented using 6 electronic databases (MEDLINE, Embase, Joanna Briggs Institute [JBI], the Cumulative Index to Nursing and Allied Health [CINAHL], the Association for Computing Machinery [ACM] library, and the IT database Institute of Electrical and Electronics Engineers [IEEE]). We also conducted gray literature searches to identify any papers that may have been missed through the search.

Screening and Study Selection
All search results were stored in an Endnote library (Endnote X9.3.3, Clarivate), and duplicates were removed. Next, the Endnote library was uploaded to Covidence (Veritas Health Innovation Ltd), and additional duplicates were removed. The PRISMA-ScR model was used to screen and select studies. Title and abstract screening and full-text screening were conducted based on the inclusion criteria. Two reviewers (authors RH and AT) performed the source selection independently. Any disagreements were discussed between the 2 reviewers, and if the conflicts were not resolved, further discussion with a third reviewer (author RR) was undertaken.

Data Exaction and Presentation of Results
Two authors conducted data extraction independently (authors RH and AT), with consensus provided by a third reviewer (RR). The data were extracted using predeveloped data extraction tables. The extracted results were descriptively mapped in tables and a narrative summary.

Consultation Exercise
One author (PR) presented an overview of the results of the scoping review to an established youth advisory group, which includes 16 adolescents aged 13–18 years, residing in New South Wales, Australia (Health Advisory Panel for Youth at the University of Sydney [HAPYUS]). The youth advisory group was recruited via social media advertising and went through a competitive selection process. They serve a 12-month term on the panel, providing their input to several adolescent research projects [30,31]. The results were presented to the youth advisory group to gain valuable insights into issues relating to the results that the scoping review alone would not have alerted the research team to. After presentation of the scoping review, 2 members of the youth advisory group volunteered to lead a statement on behalf of the group, included in the Results section, relating to considerations for researchers or developers working in this area. This statement was written by HAPYUS in their own words.

Ethical Considerations
Ethics approval was not required. The adolescents who took part in the consultation were considered members of our research team.

Results

Study Selection
The search identified 5558 papers that were imported for screening, and 85 (1.5%) duplicates were removed. After title and abstract screening, 5383 (98.4%) of 5473 papers were excluded. The remaining 90 (1.6%) full-text papers were screened, and 86 (95.6%) papers were excluded. Overall, a total of 4 (4.4%) relevant papers were identified through database searching. One additional paper was discovered through gray literature searching (Figure 1). The 5 studies were conducted in different countries: Korea, India, Finland, Switzerland, and Belgium. Among the 5 studies, 3 (60.0%) interventional studies were identified, 2 (66.7%) of which were randomized controlled trials (RCTs) and 1 (33.3%) was a pre-post study. In addition,
1 (20.0%) of the 5 studies was an exploratory analysis as a subset of an RCT and 1 (20.0%) was a mixed methods pilot study. A narrative summary of the results of the included studies and characteristics of chatbots is presented in Tables 1 and 2, respectively.

Figure 1. PRISMA-ScR flow diagram. PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for scoping reviews.
Table 1. Key characteristics of included studies.

<table>
<thead>
<tr>
<th>First author, year, country</th>
<th>Study design</th>
<th>Recruitment</th>
<th>Participants, N</th>
<th>Sex</th>
<th>Age range (years)</th>
<th>Aim</th>
<th>Use of codesign in chatbot development</th>
<th>Dropout</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lee, 2017, Korea [32]</td>
<td>Pre-post intervention</td>
<td>Students from 2 same-sex high schools in Seoul, Korea</td>
<td>33</td>
<td>Female: n=24, 72.7%; Male: n=9, 27.3%</td>
<td>16-18</td>
<td>To test the feasibility of a mobile app Diet-A and examine whether Diet-A could be used to monitor dietary intake among adolescents</td>
<td>N/A⁴</td>
<td>N/A</td>
</tr>
<tr>
<td>Padman, 2017, India [33]</td>
<td>Exploratory analysis</td>
<td>Students from 3 middle schools in urban India recruited for an RCT and de-identified participants from the RCT recruited in the explanatory analysis</td>
<td>14</td>
<td>Female: n=7, 50.0%; Male: n=7, 50.0%</td>
<td>10-11</td>
<td>To analyze game telemetry to understand user interactions from playing Fooya! and provide new insight for designing interventions via games to improve pediatric overweight and obesity rates</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Pyky, 2017, Finland [34]</td>
<td>RCT⁵</td>
<td>Males who were conscripted for military service in Finland</td>
<td>496</td>
<td>Male: n=100, 100.0%</td>
<td>Mean 17.8</td>
<td>To assess whether a tailored mobile physical activity intervention can improve life satisfaction and self-rated health among young adolescent men</td>
<td>16-20-year-old males involved in the design, development, and testing of the mobile service</td>
<td>Lost to follow-up: n=151, 30.4%; Controls: n=167, 33.7%; Intervention: n=135, 27.2%</td>
</tr>
<tr>
<td>Stasinaki, 2021, Switzerland [35]</td>
<td>RCT</td>
<td>Children’s Hospital of Eastern Switzerland (specialized childhood obesity management center)</td>
<td>31</td>
<td>Female: n=13, 41.9%; Male: n=18, 58.1%</td>
<td>10-18</td>
<td>To assess whether PathMate2 can improve the BMI (kg/m²), physical capacities, and stress parameters in adolescents with obesity, under the supervision of pediatric obesity experts</td>
<td>N/A</td>
<td>Lost to follow-up: 0.1%</td>
</tr>
<tr>
<td>Maenhout, 2021, Belgium [36]</td>
<td>Mixed methods pilot study</td>
<td>Flemish secondary schools</td>
<td>Phase 1: 36; Phase 2: 6; Phase 3: 81</td>
<td>Female: n=29, 80.6%; Male: n=7, 19.4%; Phase 1: Female: n=6, 100.0%; Phase 3: N/A</td>
<td>12-15</td>
<td>To assess the feasibility and engagement of a chatbot prototype among adolescents to promote healthy behaviors</td>
<td>Phase 1: focus groups to inform the development of the chatbot prototype, including content and design; Phase 2: pretest of the prototype</td>
<td>Phase 3: quit after receiving a wrong answer from the chatbot: n=61, 66.7%</td>
</tr>
</tbody>
</table>

⁴N/A: not applicable. ⁵RCT: randomized controlled trial.

Table 2. Summary of chatbots.

<table>
<thead>
<tr>
<th>First author, year, country</th>
<th>Chatbot name</th>
<th>Intervention delivery</th>
<th>Conversational agent</th>
<th>Gamification</th>
<th>Personalized feedback</th>
<th>Monitored behavior change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lee, 2017, Korea [32]</td>
<td>Diet-A</td>
<td>Mobile app</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Padman, 2017, India [33]</td>
<td>Fooya!</td>
<td>Mobile app</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Pyky, 2017, Finland [34]</td>
<td>MOPOrtal</td>
<td>Mobile service</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Stasinaki, 2021, Switzerland [35]</td>
<td>PathMate2</td>
<td>Mobile app</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Maenhout, 2021, Belgium [36]</td>
<td>Self-regulation app</td>
<td>Mobile app</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Overview of Included Studies

Studies recruited adolescents aged 10-19 years. Of the 5 studies, 4 (80.0%) had small sample sizes with varying distributions of male and female participants. In addition, 1 (20.0%) study had an even distribution of males and females [33], 2 (40.0%) studies had more than 70.0% female participants [32,36], and 2 (40.0%) studies had predominantly (58.0% and 100.0%, respectively) male participants [34,35]. An overview of the included studies is provided in Table 1, including key characteristics of the included studies (eg, authors, year of publication, country, aim, study type, participant characteristics). In the included studies, 1 (20.0%) study included adolescents with overweight or obesity recruited from a hospital setting [35]. The other 4 (80.0%) studies included participants who were otherwise healthy [32-34,36]. In addition, 3 (60.0%) studies were conducted in school settings, of which 2 (66.7%) were conducted in high schools and 1 (33.3%) in middle schools [32,33,36]. Furthermore, 1 (20.0%) study recruited only males eligible for military conscription [34].

Summary of Chatbots

The 5 chatbots were supported by mobile apps (n=4, 80.0%) or web applications delivered via mobile devices (n=1, 20.0%). The 5 chatbots were different in their delivery. The chatbots used a combination of 4 features, namely a conversational agent (n=4, 80.0%), gamification (n=3, 60.0%), personalized feedback (n=4, 80.0%), and monitoring of behavior change (n=4, 80.0%). An overall summary of the chatbots is provided in Table 2, including the characteristics of the intervention in more detail (chatbot details, intervention details, outcomes and key findings that relate to the scoping review question). In the 5 studies, chatbots were used in different ways to improve adolescents’ nutrition and physical activity behaviors. Nutrition was the focus of 2 (40.0%) studies, in which chatbots targeted nutrition intake and food choice [32,33]. Physical activity was the focus of 2 (40.0%) studies, in which chatbots targeted physical activity, physical capacity, and the BMI [34,35]. Finally, 1 (20.0%) study had a chatbot that targeted both nutrition and physical activity behaviors [36]. Each intervention targeted nutrition and physical activity differently. Diet-A used a mobile app where the participants recorded their dietary intake and provided real-time, personalized feedback on their diet [32]. Fooya! was an interactive mobile game and AI robot that aimed to influence healthy food choices [33]. The 2 chatbots targeting physical activity had unique features in their delivery to help participants achieve their goals. PathMate2 was a virtual health coach [35], and MOPOrtal was a web-based interface with a combined mixed-reality game [34]. The self-regulation app that targeted physical activity had unique features in their delivery to help participants achieve their goals. PathMate2 was still being used by just over half of the participants (51.0%) at 6 months. The average app usage rate was 71.5%, and the average adherence rate was 57.2% during the intervention [35]. Finally, for the self-regulation app, 74.1% (60/81) of participants used the chatbot during the pilot; however, two-thirds of these participants quit and did not ask any further questions if the chatbot gave a wrong answer [36].

Summary of Health Outcomes

Of the 5 studies, 3 (60.0%) studies recorded and analyzed participants’ health-related characteristics at baseline and after the intervention, with the length of the interventions ranging from 3 to 6 months [32,34,35], and 1 (20.0%) study had an additional 6-month maintenance phase to measure sustained changes [35]. The Diet-A intervention used the CAN-Pro 4.0 program to assess nutrient intake through 24-hour recalls pre- and postintervention. This study found that participants had a significant reduction in sodium and calcium intake and an increase in fruit and vegetable consumption. However, there was no improvement in overall diet among the participants following the intervention [32]. MOPOrtal measured daily minutes of physical activity through a physical activity monitor and collected height and weight to calculate the BMI. It demonstrated a limited increase in physical activity and increased mean weight in both intervention and control groups. Only those men with low life satisfaction and poor self-rated health at baseline were associated with improved satisfaction postintervention [34]. Finally, PathMate2 measured the BMI-SDS (where SDS refers to the standard deviation score) and other anthropometric measures and found that participants can improve physical capacity, increase muscle mass, and reduce body fat percentage following use of the intervention, but there was no sustained significant change in the BMI-SDS [35]. The other 2 (40.0%) studies did not measure any health-related characteristics. A full summary of outcomes is provided in Multimedia Appendix 2.

Summary of Chatbot Development

Of the 5 chatbots, 4 (80.0%) used text-based mobile apps yet were developed in different ways, including based on health databases, transtheoretical models, scientific evidence, and the person-based approach (PBA) [32,34-36]. The mobile app Diet-A, developed by Lee et al [32], is a self-monitoring app to help participants record their diet and offers real-time feedback and disease prevention information based on dietary reference intakes for Koreans. The feedback and disease prevention information were built under 3 health and prevention information were built under 3 health and prevention information were built under 3 health and prevention information were built under 3 health and medical knowledge and awareness of healthy food, but engagement decreased throughout the game [33]. In the MOPOrtal intervention, there were low overall intervention effects, except in participants who reported poorer health at baseline. No other data on feasibility or acceptability were reported [34]. PathMate2 was still being used by just over half of the participants (51.0%) at 6 months. The average app usage rate was 71.5%, and the average adherence rate was 57.2% during the intervention [35]. Finally, for the self-regulation app, 74.1% (60/81) of participants used the chatbot during the pilot; however, two-thirds of these participants quit and did not ask any further questions if the chatbot gave a wrong answer [36].

Participants reported that it was burdensome and 85.7% (18/21) reported that they sometimes forgot to record their diet [32]. In the Fooya! mobile app, participants gained knowledge and awareness of healthy food, but engagement decreased throughout the game [33]. In the MOPOrtal intervention, there were low overall intervention effects, except in participants who reported poorer health at baseline. No other data on feasibility or acceptability were reported [34]. PathMate2 was still being used by just over half of the participants (51.0%) at 6 months. The average app usage rate was 71.5%, and the average adherence rate was 57.2% during the intervention [35]. Finally, for the self-regulation app, 74.1% (60/81) of participants used the chatbot during the pilot; however, two-thirds of these participants quit and did not ask any further questions if the chatbot gave a wrong answer [36].
13-18-year-olds. The messages delivered were based on the transtheoretical model of behavior change. The given message was different at each intervention stage to match the process of change theorized and provide the most appropriate information to the participants. The health information was based on the reviewed scientific evidence [34]. The PathMate2 mobile app included a conversational agent as a virtual coach and was developed with MobileCoach open source software. This agent can chat with participants and encourage them to achieve the challenge of staying healthy through physical activity according to Swiss physical activity guidelines. PathMate2 aimed to support behavior change using goal setting, self-monitoring, stimulus control, and behavioral contracting to support a healthy lifestyle [35]. Finally, Maenhout et al [36] used PBA to ensure the needs and perspectives of the end user were embedded in the guiding principles of the chatbot, and therefore, health information delivered was not based on any guidelines but rather was based on content adolescents wished to receive. Dialogflow software was used to develop the intervention, and behavior change was promoted using the Health Action Process Approach model [36]. Moreover, there was 1 (20.0%) study that examined a virtual reality–based mobile game that was supported by AI (food robot), which was different from the other 4 (80.0%) chatbots and used personalized behavior reinforcement to increase awareness and self-efficacy [33]. Only 2 (40.0%) of 5 studies used any codesign with the end user, and no studies involved parents or caregivers in the intervention development.

Youth Consultation
The youth consultation led to the statement seen in Textbox 1. In brief, adolescents had concerns around (1) information the chatbots delivered being misleading or harmful and (2) ethical concerns around the privacy of data collected and misunderstanding of individual circumstances that may provide inaccurate health advice.

Textbox 1. Youth statement in their own words.

Chatbots have great potential in the field of health promotion, particularly in areas that encompass physical activity and nutrition. However, there are many factors that must be considered before they are implemented in such a field. The extensive growth and use of social media and the sharing of public information have seen society enter a world of fake, or rather, misleading information. This has created an environment where it is hard to navigate what is the truth and what is harmful. Therefore, any information that the chatbots release must be highly regulated and fact-checked before being released. So many misleading and often harmful nutritional messages are put out to audiences that result in body dysmorphia, decreases in self-esteem, and eating disorders. The information used must be phrased in a manner that is not triggering nor encouraging such poor habits. To increase their acceptance in the wider population, the chatbot should be associated with a brand or source that already has a “trusted” label. This would make audiences more likely to engage with it.

The ethical concerns of chatbots for uses in health promotion can be divided into 2 main categories: the potential for chatbots to exploit young people for commercial gain and the potential for chatbots to cause harm to young people through the provision of inaccurate health advice. There are several ways in which chatbots could exploit young people for commercial gain. Chatbots could be used to sell young people’s personal data to third parties or to generate targeted advertising based on young peoples’ health conditions. Chatbots could also be used to upsell young people on expensive treatments, exercise programs, or supplements. To minimize the risk of chatbots exploiting young people for commercial gain, it is important to ensure that chatbots are transparent about how they will use any personal data that they collect. Young people should also be given the option to opt out of any data collection or advertising. There is also a risk that chatbots could cause harm to young people through the provision of inaccurate health advice. This could happen if chatbots are not based on credible health sources or if they are not able to properly understand young people’s individual circumstances. To minimize the risk of chatbots causing harm to young people, it is important to ensure that chatbots are only used as a supplement to, and not a replacement for, health advice from a qualified health care professional.

Discussion
Principal Findings
This systematic scoping review evaluating chatbots in promoting nutrition and physical activity behaviors in adolescent populations is an emerging and underresearched field. The 5 published studies found insufficient evidence for the acceptability and feasibility of chatbots. Only 2 of the 5 included studies found adolescents were satisfied with the chatbot used in the intervention [32,33]. The chatbots demonstrated modest efficacy in improving adolescents’ nutrition, physical activity behaviors, and knowledge. The chatbots were used within mobile apps or mobile services with differing design features, including conversational agents, gamification, personalized feedback, and monitoring of behavior change. Adolescents from the youth advisory group presented unique insights into the use of chatbots in nutrition and physical activity interventions, including ethical concerns and the use of false or misleading information, which was not otherwise identified in the published literature. Taking these findings together, this review found that there is limited evidence for the feasibility and acceptability of chatbots in promoting nutrition and physical activity behaviors. Therefore, together with our youth advisory group, we propose suggestions for improved chatbot development and research study design.

Comparison With Existing Literature
To the best of our knowledge, this is the first systematic scoping review of chatbots in promoting nutrition and physical activity behaviors in adolescent populations. Chatbots have been broadly used in chronic disease prevention and management. A systematic review conducted by Laranjo et al [37] demonstrated that conversational agents are most commonly used in mental health management, resulting in reduced depression symptoms, improved narrative skills scores in autism, and suicide prevention [37]. This review also highlighted conversational agents (1) supporting patients with type 2 diabetes for physical activity and diet behavior change and self-management practice and (2) supporting clinicians and hypertension patients in telemonitoring and data collection. In this review, 12 of the 14 studies reported user experience. Dissimilar to our findings, most reported high overall satisfaction. Of the 2 studies that included adolescent participants, the chatbot designed for self-management of a specific condition (asthma) [38] had a
higher overall satisfaction compared to the chatbot designed for education (sexual health and substance abuse) [39]. In our review, participants reported greater satisfaction with chatbots for self-monitoring of food consumption and dietary intake. However, most participants reported that chatbots were often not easy to use and sometimes forgot to record their dietary intake [32]. Consequently, participants tended to underreport their dietary intake using the app in comparison to other validated dietary recall methods. This may explain why studies in our review, which were focused on prevention and risk factor modification, not chronic disease self-management, had lower overall satisfaction. This also demonstrates the need to focus on design features.

There are other studies focusing on the feasibility of chatbots used in adolescents but not limited to nutrition and physical activity behavior change. A chatbot called Tess (X2 AI) using AI was found to be an engaging and feasible approach to support weight management and counseling in adolescents and children [40]. Participants reported Tess to be useful 96.0% of the time. The high level of satisfaction compared to the studies included in our review may be explained by the different length of conversations participants can have with Tess. Tess can offer large amounts of message exchanges, which demonstrates high engagement, attraction, and acceptability of AI chatbots [40]. It should be noted that Tess is a commercially available service with a customizable platform where the content can be tailored for specific populations or interventions. This is unlike the chatbots evaluated in this review, which were developed by the research teams for the purpose of 1 intervention. Consequently, integrating language techniques may be useful to incorporate into the chatbot database to enhance engagement with adolescents and stimulate longer message conversations, covering topics outside of the intervention itself.

NLP may be a good choice for chatbot database design for adolescents if databases can be developed to offer small talk and noninterventional questions, in addition to the intervention. In Maenhout et al’s [36] study, adolescents found it frustrating if the chatbot misunderstood their question. A conservationist agent that uses NLP may make them feel like they are communicating with another human, which in turn may enhance engagement and the user experience [36]. In a similar study, a chatbot using NLP that was focused on improving physical activity in adults found the chatbot increased participants’ step count and self-reported physical activity. Most participants scored the chatbot as OK (78.8%), and one-third of the participants were interested in continuing using the chatbot following the study [41]. For NLP to be successful, it is vital to engage adolescents throughout the database design process to develop the database with youth-oriented language and enhance the feeling of communicating with another human. A scoping review by Kramer et al [42] found that conversational agents for coaching people in a healthy lifestyle were often designed for the end user rather than with the end user. In this review, only 2 of the 5 chatbots incorporated any kind of codesign with adolescents in the development of the chatbot intervention, which may explain the low satisfaction with the chatbots as they were not designed with the end user, and therefore do not meet the adolescents’ needs.

The youth consultation uncovered insights into the use of chatbots for nutrition and physical activity interventions that were not identified in the published literature. One of the suggestions raised by adolescents was to have the chatbot associated with a brand. In a previous study, adolescents identified that the most helpful lifestyle health information online comes from a credible and reliable source [43]. Adolescents are highly brand conscious [44], and therefore, having the chatbot associated with a brand may increase their trust in the information that is being presented to them. Another insight raised by adolescents was around the provision of inaccurate health advice that may cause harm. To counter this, appropriate monitoring of chatbot conversation logs is vital in future studies to ensure chatbots do not deviate and provide incorrect information to adolescents. Conversation logs must also be monitored to ensure any self-disclosure from the adolescent to the chatbot is communicated and actioned accordingly. In the studies included in this review, there was no potential for chatbots to provide incorrect information as none of the conversational agents used AI to provide responses. Only 1 of the studies in this review applied monitoring of conversation logs, yet it was to assess the feasibility and not for safety. For chatbots to be both safe and effective in the future, researchers and developers must work together to obtain information about adolescents and their individual situation and then tailor accurate health information that is best suited to their needs. Furthermore, safeguards need to be in place to ensure the safety of adolescents while using chatbots for health promotion interventions [45], especially if future chatbots are developed using AI. Rigorous beta-testing of the intervention should occur before being implemented to ensure that interventions are relevant, appealing, functional, stable, and useful [46]. In addition, exposure time to the chatbots must also be considered in future interventions to ensure that adolescents do not increase their screen time beyond the recommended guidelines.

Limitations
This scoping review demonstrates the limited published literature on chatbots used in the adolescent population for nutrition and physical activity behavior promotion. It must be noted that there are some limitations to this research. First, not all studies provided data on the feasibility and acceptability of the chatbots, which is crucial to understanding barriers and enablers to implementing such an intervention on a wider scale. Second, none of the studies included in this review that included a conversational agent used AI. Chatbots based on AI are trained to respond to queries based on texts to which they are exposed; therefore, the training of AI chatbots could not be assessed within the scope of this review. Next, we only included peer-reviewed published studies. There is the potential of other studies that would otherwise fit the criteria of this review.

Conclusion
Limited research is available on the use of chatbots in adolescent nutrition and physical activity interventions, finding insufficient...
evidence for the acceptability and feasibility of chatbots in the adolescent population and only minor improvements in health-related outcomes due to the interventions. Similarly, adolescent consultation identified important issues relating to the design features that were not mentioned in the published literature. Researchers and developers should consider codesigning chatbots with adolescents to ensure that they are feasible and acceptable to an adolescent population.

Acknowledgments
A special acknowledgment to Fulin Yan and Alexi Cross for leading the youth consultation and the entire Health Advisory Panel for Youth at the University of Sydney 2021/22 (HAPYUS 2021/22), without whom this research would not be possible (Radhika Valanju, Meera Barani, Dominik Mautner, Inemelya Al Hadaya, Melani Gunawardana, Ava Lambie, Emily McMahon, Arnav Narula, Bowen Ren, Dominique R, Aviral Sharda, Alexander Sinnett, and Azman Tanvir).

Authors' Contributions
Conceptualization was handled by RH, SRP, and RR; methodology by RH, SRP, and RR; investigation by RH, SW, and AT; writing—original draft preparation by RH and AT; writing—review and editing by RH, AT, SW, SRP, and RR; and supervision by SRP and RR. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File, 16 KB - humanfactors_v10i1e43227_app1.docx ]

Multimedia Appendix 2
Characteristic of chatbots and interventions.
[DOCX File, 20 KB - humanfactors_v10i1e43227_app2.docx ]

References
1. Adolescent health. World Health Organization. URL: https://www.who.int/health-topics/adolescent-health#tab=tab_1 [accessed 2022-09-28]


### Abbreviations  

- **AI**: artificial intelligence  
- **HAPYUS**: Health Advisory Panel for Youth at the University of Sydney  
- **NLP**: natural language processing  
- **PBA**: person-based approach  
- **PRISMA-ScR**: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for scoping reviews  
- **RCT**: randomized controlled trial  
- **WHO**: World Health Organization  

---  

Edited by A Kashmirak; submitted 04.10.22; peer-reviewed by M Vajravelu, T Baranowski; comments to author 26.02.23; revised version received 15.03.23; accepted 13.04.23; published 05.05.23.  

Please cite as:  
Han R, Todd A, Wardak S, Partridge SR, Raeside R  
Feasibility and Acceptability of Chatbots for Nutrition and Physical Activity Health Promotion Among Adolescents: Systematic Scoping Review With Adolescent Consultation  
JMIR Hum Factors 2023;10:e43227  
URL: https://humanfactors.jmir.org/2023/1/e43227  
doi:10.2196/43227  
PMID:37145858  

©Rui Han, Allyson Todd, Sara Wardak, Stephanie R Partridge, Rebecca Raeside. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 05.05.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Factors That Affect Knowledge-Sharing Behaviors in Medical Imaging Departments in Cancer Centers: Systematic Review

Maryam Almashmoum1,2, BSc, MSc; James Cunningham1, MSc, PhD; Ohoud Alkhaldi1,3, BSc, MSc; John Anisworth1, MSc, PhD

1Division of Informatics, Imaging and Data Sciences, School of Health Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, United Kingdom
2Nuclear Medicine Department, Faisal Sultan Bin Eissa, Kuwait Cancer Control Center, Kuwait City, Kuwait
3Health Information Management and Technology Department, College of Public Health, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia

Corresponding Author:
Maryam Almashmoum, BSc, MSc
Division of Informatics, Imaging and Data Sciences
School of Health Sciences
Faculty of Biology, Medicine and Health, The University of Manchester
Vaughan House, Portsmouth St
Manchester, M13 9GB
United Kingdom
Phone: 44 07949062010
Email: maryam.almashmoum@postgrad.manchester.ac.uk

Abstract

Background: Knowledge management plays a significant role in health care institutions. It consists of 4 processes: knowledge creation, knowledge capture, knowledge sharing, and knowledge application. The success of health care institutions relies on effective knowledge sharing among health care professionals, so the facilitators and barriers to knowledge sharing must be identified and understood. Medical imaging departments play a key role in cancer centers. Therefore, an understanding of the factors that affect knowledge sharing in medical imaging departments should be sought to increase patient outcomes and reduce medical errors.

Objective: The purpose of this systematic review was to identify the facilitators and barriers that affect knowledge-sharing behaviors in medical imaging departments and identify the differences between medical imaging departments in general hospitals and cancer centers.

Methods: We performed a systematic search in PubMed Central, EBSCOhost (CINAHL), Ovid MEDLINE, Ovid Embase, Elsevier (Scopus), ProQuest, and Clarivate (Web of Science) in December 2021. Relevant articles were identified by examining the titles and abstracts. In total, 2 reviewers independently screened the full texts of relevant papers according to the inclusion and exclusion criteria. We included qualitative, quantitative, and mixed methods studies that investigated the facilitators and barriers that affect knowledge sharing. We used the Mixed Methods Appraisal Tool to assess the quality of the included articles and narrative synthesis to report the results.

Results: A total of 49 articles were selected for the full in-depth analysis, and 38 (78%) studies were included in the final review, with 1 article added from other selected databases. There were 31 facilitators and 10 barriers identified that affected knowledge-sharing practices in medical imaging departments. These facilitators were divided according to their characteristics into 3 categories: individual, departmental, and technological facilitators. The barriers that hindered knowledge sharing were divided into 4 categories: financial, administrative, technological, and geographical barriers.

Conclusions: This review highlighted the factors that influenced knowledge-sharing practices in medical imaging departments in cancer centers and general hospitals. In terms of the facilitators and barriers to knowledge sharing, this study shows that these are the same in medical imaging departments, whether in general hospitals or cancer centers. Our findings can be used as guidelines for medical imaging departments to support knowledge-sharing frameworks and enhance knowledge sharing by understanding the facilitators and barriers.

(JMIR Hum Factors 2023;10:e44327) doi:10.2196/44327
Introduction

Background

Knowledge is an important element in the success of many institutions. It allows institutions to gain competitive advantages and aids institutional survival [1,2]. To maintain these benefits, many institutions use massive resources to implement knowledge management systems and encourage knowledge sharing among the health care professionals within those institutions. Moreover, it is considered one of the main assets in health care institutions that can be used to achieve the best patient outcomes [3]. Davenport and Prusak [4] defined knowledge as “a fluid mix of framed experience, value, contextual information, and expert insights that provides a framework for evaluating and incorporating new experiences and information.” Polanyi [5] classified knowledge into 2 categories: tacit and explicit. Explicit knowledge is easy to codify and generate in tangible forms, for example, documents, manuals, and policies [6]. In contrast, tacit knowledge is the knowledge that exists in human minds and individuals’ experiences. These experiences might be revealed through interactions with health care professionals within the workplace [7].

Health care institutions are building their own knowledge management systems to enhance the use of their own knowledge [8]. The success of any knowledge management program depends on communication among health care professionals in general and sharing knowledge among them in particular [7-9]. Health care institutions have a knowledge-sharing culture by changing health care professionals’ attitudes and behavior [8]. The concept of knowledge management is related to sharing ideas, thoughts, and experiences among health care professionals to improve health care settings [8]. In contrast, effective knowledge sharing among health care professionals depends on several facilitators, and barriers affect knowledge-sharing practices. The lack of awareness of knowledge sharing in health care institutions is the main barrier to establishing knowledge-sharing practices [9]. Moreover, according to Tetroe et al [10], knowledge sharing is essential to health care, whether in the public or private sector, and it can offer greater responsibility in health planning, making decisions, and delivering several services. From a health care point of view, knowledge sharing is a crucial instrument to ensure that the correct information gets to the right person and is used for specific purposes in the right environment at the right time [11].

Knowledge sharing among health care professionals in medical imaging departments in cancer centers plays a vital role in cancer survivorship by promoting communication among health care professionals, thus enabling them to understand cases in depth with input from professionals across different disciplines and facilitating the best interpretation of results [12]. Each cancer center has unique policies to enhance knowledge sharing among their health care professionals in a particular way without affecting patient outcomes. These policies are controlled at several points to protect patients’ privacy and help them receive appropriate treatment and make correct decisions regarding their case. Furthermore, in 2016, the National Radiotherapy Advisory Group strongly recommended that National Health Service radiography services should be increased to approximately 90% to keep up with the aging population and earlier detection of cancer cases [13]. There are several actors involved in medical imaging departments, such as physicians, oncologists, radiographers, radiologists, and nuclear medicine technologists. Technologists represent the third largest group of health care professionals [14]. In addition, approximately 60% of the health care professional workers comprise allied health professionals in the United States [15]. These allied health professionals play a crucial role in medical imaging departments and have gained plenty of knowledge in their field, either theoretical or practical, and this knowledge has to be shared among them to improve patient outcomes [16]. To improve patient care and outcomes, it is important to focus on knowledge sharing among health care professionals [17]. In medical imaging departments, knowledge sharing is complex as it involves visual patterns created using plain-text annotations and images. Therefore, knowledge sharing in medical imaging departments requires a system base to share these images, for example, the picture archiving and communication system (PACS) [18].

On the basis of previous studies on knowledge sharing, the factors that affect knowledge sharing were divided into categories: facilitators and barriers. The facilitators that enhance knowledge sharing in health care institutions were classified into 3 categories: individual, departmental, and technological facilitators. The barriers that hinder knowledge sharing were divided into 4 categories: financial, administrative, technological, and geographical barriers [19-32].

Types of Knowledge in Medical Imaging Departments

Tacit Knowledge

Tacit knowledge is a vital part of human reasoning. It revolves around how humans interact with each other in their surrounding environment [33]. When explicit knowledge fails to present a full explanation of an idea, tacit knowledge can help draw a clear explanation and reach a conclusion. Moreover, it is difficult to share in its nature because of the human tendency to own their knowledge, which can give them an advantage over other peers in an institution. Tacit knowledge is exhibited in medical imaging departments as thoughts, ideas, experiences, and interpretations of results regarding specific cases [33]. Moreover, tacit knowledge is embodied in routine daily work among health care professionals everywhere, even in the hospital corridors. Furthermore, tacit knowledge is considered a lecturer’s tool, which is very important for disseminating knowledge [34]. Radiologists by nature prefer to establish contact face-to-face with each other in subgroups to share their common interests [35].
Peer-to-peer networks are considered one of the most successful ways to share tacit knowledge in medical imaging departments [33]. Storytelling is a practical way to share knowledge among health care professionals in these departments. It takes place during a medical diagnosis [34]. Teamwork meetings and conferences, whether physically or digitally, allow tacit knowledge to be a dominant type of knowledge that emerges from these gatherings [34].

**Explicit Knowledge**

Explicit knowledge is knowledge that exists in a tangible form. It is easy to generate in different forms, such as documents and policies. It exists in medical imaging departments as various documents containing information such as policies, procedure manuals, hospital protocols, and quality assurance documents for monthly records [35]. Any health care professional who takes on a role in the medical imaging department has a responsibility to know these documents and how to record them monthly. These documents are stored in an accessible place to be easily referred back to at any time [36]. Moreover, these documents are stored either manually or electronically to avoid losing them under any circumstances. These documents are updated annually when necessary [36].

Sharing explicit knowledge occurs in the medical imaging department during its workday by sharing circulars, patient requests, medical imaging, and quality control for the machines monthly.

**Textbox 1. Population, intervention, control, outcome, and study design strategy.**

**Population**
- The population of interest was any health care professionals who were working in the medical imaging departments in cancer centers or not, with the included studies reporting knowledge sharing among them (radiographers, technologists, nuclear medicine specialists, physicians, practitioners, radiologists, and nurses).

**Intervention**
- This included knowledge-sharing tools, mechanisms, and procedures that enhance knowledge-sharing practices.

**Control**
- The included studies identified facilitators that enhance knowledge-sharing practices. Moreover, the studies investigated the barriers that hinder knowledge-sharing behaviors.

**Outcome**
- The general outcome of the studies was to enhance knowledge sharing among health care professionals by identifying facilitators of and barriers to knowledge-sharing practices to improve patient outcomes and services and reduce medical mistakes.

**Study design**
- The study designs involved finding the facilitators and barriers that affect knowledge-sharing practices among health care professionals in medical imaging departments in general and particularly in cancer centers.

As a result, the following research questions were addressed in this systematic review: (1) What are the facilitators of knowledge sharing among health care professionals in medical imaging departments in general and in cancer centers in particular? (2) What are the barriers that hinder knowledge-sharing practices among health care professionals in medical imaging departments in general and in cancer centers in particular? and (3) What are the differences in factors between medical imaging departments in general hospitals and cancer centers?

**Search Strategy and Sources of Information**

We searched 7 databases in December 2021: PubMed Central, EBSCOhost (CINAHL), Ovid MEDLINE, Ovid Embase, Elsevier (Scopus), ProQuest, and Clarivate (Web of Science). From ProQuest, we used 8 databases: ProQuest One Academic, ProQuest Central, Health and Medical Collection, Nursing and
Allied Health Database, Healthcare Administration Database, Public Health Database, Consumer Health Database, and Materials Science Collection. The specific reason for choosing these databases was their relationship with health care institutions. The search terms were designed to capture factors that affect knowledge-sharing practices among health care professionals in medical imaging departments in cancer centers. Medical Subject Heading terms were used with the Boolean operators AND and OR to enhance the search strategy by locating the relevant studies. The search strategies for all the databases are presented in Multimedia Appendix 1.

Eligibility Criteria
The inclusion and exclusion criteria were formulated based on the main objective of the thesis to answer the research questions. Articles were eligible if they met the following criteria: (1) studies that examined knowledge-sharing practices; (2) studies published within the last 20 years; (3) studies conducted in medical imaging departments in cancer centers; (4) studies that investigated knowledge-sharing facilitators and barriers within medical imaging departments in general and in cancer centers in particular; and (5) qualitative, quantitative, and mixed methods designs.

Articles were excluded if they met the following criteria: (1) studies published in a language other than English; (2) meeting reports, keynotes, abstracts, books, and presentations; and (3) studies related to knowledge sharing between health care professionals and patients.

Selection Process
All articles identified from the database searches were exported to EndNote Online (Clarivate Analytics), which was used for screening and eliminating any duplicates. In total, 2 reviewers (MA and OA) independently screened the titles and abstracts of all the studies. To determine whether an article should be examined in depth, the 2 reviewers assessed the article for eligibility based on the inclusion and exclusion criteria. All disagreements were resolved through discussion to make the final decision. Figure 1 shows the details of the exclusion and inclusion criteria.

Data Collection Process
After the selection of the final studies between the 2 reviewers, the primary reviewer performed the data extraction. The data extracted from the studies included the following: names of the authors, year of publication, country, sample size, facilitators, barriers, quality of the study (based on the strong evidence of the facilitators and barriers), and main findings.
Quality Appraisal: Risk of Bias of the Included Studies

Following the final selection of studies, the risk of bias was assessed using the Mixed Methods Appraisal Tool (MMAT) [38]. It is used for evaluation in reviews that include qualitative, quantitative, and mixed methods studies. According to the MMAT tool, assigning a single score based on the assessment is not recommended [38]. On the basis of a previous study, we used a specific statistical strategy [39] to assess the quality of each study to justify the final decision based on the inclusion and exclusion criteria. Depending on the number of criteria met, the studies were classified as high, medium, or low quality. A study was rated as high quality if all 5 MMAT criteria were met, as medium quality if 3 or 4 criteria were met, and as low quality if <2 criteria were met [39].

Data Synthesis

This review used narrative synthesis to summarize the evidence from the final studies that were included. Narrative synthesis is useful and appropriate as this study included qualitative, quantitative, and mixed methods findings [40].

Results

Study Selection

The electronic search retrieved 2708 study records from 7 databases (n=62, 2.29% from PubMed Central; n=15, 0.55% from EBSCOhost [CINAHL]; n=5, 0.18% from Ovid MEDLINE; n=21, 0.78% from Ovid Embase; n=620, 22.9% from Elsevier [Scopus]; n=1631, 60.23% from ProQuest; and n=354, 13.07% from Clarivate [Web of Science]). After duplicates, which were 1.51% (41/2708) of the articles, and other articles (124/2708, 4.58%) were removed manually, of the 2708 records, 2543 (93.91%) studies remained that were assessed for title and abstract screening. In total, 1.93% (n=354) of the studies were eligible for full-text screening, and the final number of studies included in the review was 38. In addition, 1 article was added from selected databases. A total of 39 studies were included in the review. The process and reasons for selecting these studies are shown in Figure 1.

Characteristics of the Included Studies

Multimedia Appendix 2 [1, 12, 18, 33-36, 41-72] summarizes the characteristics of the included studies. The 39 studies included in the review were from different areas of the world: 5 (13%) were from the United States; 7 (18%) were from Canada; 6 (15%) were from Australia; 2 (5%) were from the Netherlands; 2 (5%) were from the United Kingdom; 2 (5%) were from Saudi Arabia; 2 (5%) were from Germany; and 13 (33%) were from Taiwan, Ireland, Italy, Kenya, India, Kuwait, South Africa, Sweden, France, Iran, Brazil, Finland, and Norway. Most of the included studies (26/39, 67%) presented the factors and facilitators that affect knowledge sharing among health care professionals in medical imaging departments in cancer centers. Of the 39 studies, 26 (67%) were conducted in medical imaging departments in cancer centers, whereas 13 (33%) were conducted in medical imaging departments without mentioning whether they were in cancer centers. Of the 39 studies, 21 (54%) used qualitative methods (interviews, semistructured interviews, and case studies), 14 (36%) used quantitative methods (either surveys or questionnaires), and 4 (10%) used a mixed methods approach.

The quality of the articles is shown in Multimedia Appendix 2 [1, 12, 18, 33-36, 41-72]. There were a few articles (2/39, 5%) considered weak as they related to knowledge management in general, types of knowledge, and how it is documented without any evidence of factors that affect knowledge sharing [33, 34]. Therefore, those articles need to be documented in this study. For example, Barb et al [33] suggested that a successful way to share tacit knowledge is through peer-to-peer networks, whereas Zucchermaglio and Alby [34] argued that storytelling is a practical way to share tacit knowledge in medical imaging departments. Therefore, tacit knowledge has become a dominant type of knowledge in medical imaging departments [34].

Quality of the Included Studies

Multimedia Appendix 3 [1, 12, 18, 33-36, 41-72] highlights the results of the quality assessment of the included studies. Of the 39 studies, 26 (67%) were rated as high quality as they met all 5 MMAT criteria (11/26, 42% qualitative; 12/26, 46% quantitative; and 3/26, 12% mixed methods), 11 (28%) were rated as medium quality as they met 3 or 4 of the MMAT criteria (9/11, 82% qualitative; 1/11, 9% quantitative; and 1/11, 9% mixed methods), and 2 (5%) qualitative studies were evaluated as having a low quality as they met <2 of the MMAT criteria. This review was exploratory in nature. Therefore, we decided not to exclude these studies from the final review based on the low quality regarding the MMAT criteria.

Synthesis of the Results

Overview

The facilitators and barriers that affect knowledge-sharing behaviors in medical imaging departments are presented in Tables 1 and 2. We categorized the reported facilitators and barriers based on their apparent commonality according to the descriptions in previous studies [19-32]. These facilitators were divided according to their characteristics into 3 categories: individual, departmental, and technological facilitators. The barriers that hindered knowledge sharing were divided into 4 categories: financial, administrative, technological, and geographical barriers.
<table>
<thead>
<tr>
<th>Category and facilitator</th>
<th>Studies, n (%)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Individual facilitators</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive attitude</td>
<td>3 (8)</td>
<td>Lam et al [41]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Taba et al [42,43]</td>
</tr>
<tr>
<td>Awareness</td>
<td>1 (3)</td>
<td>Al Mashmoum and Hamade [44]</td>
</tr>
<tr>
<td>Experience</td>
<td>1 (3)</td>
<td>Taba et al [42]</td>
</tr>
<tr>
<td>Intrinsic motivation</td>
<td>5 (13)</td>
<td>Al Mashmoum and Hamade [44]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kilsondon et al [45]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Singh et al [46]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Welter et al [47]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Armoogum and Buchgeister [48]</td>
</tr>
<tr>
<td>Self-efficacy</td>
<td>2 (5)</td>
<td>Al Mashmoum and Hamade [44]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Singh et al [46]</td>
</tr>
<tr>
<td>Self-esteem</td>
<td>2 (5)</td>
<td>Taba et al [42,43]</td>
</tr>
<tr>
<td>Trust</td>
<td>5 (13)</td>
<td>Al Mashmoum and Hamade [44]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bagayogo et al [49]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fatahi et al [50]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Moilanen et al [51]</td>
</tr>
<tr>
<td>Personality</td>
<td>2 (5)</td>
<td>Al Mashmoum and Hamade [44]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Patton [52]</td>
</tr>
<tr>
<td><strong>Departmental facilitators</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Community of oncologists</td>
<td>1 (3)</td>
<td>Dicicco-Bloom and Cunningham [53]</td>
</tr>
<tr>
<td>Community of practice</td>
<td>4 (10)</td>
<td>Armoogum and Buchgeister [48]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Glicksman et al [54]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fingrut et al [55,56]</td>
</tr>
<tr>
<td>Departmental arrangements</td>
<td>1 (3)</td>
<td>Al Mashmoum and Hamade [44]</td>
</tr>
<tr>
<td>Leadership</td>
<td>6 (15)</td>
<td>Dorow et al [35]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lam et al [41]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Al Mashmoum and Hamade [44]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Moilanen et al [51]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Patton [52]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lee et al [57]</td>
</tr>
<tr>
<td>Culture</td>
<td>4 (10)</td>
<td>Patton [52]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fingrut et al [55]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Taba et al [43]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mork-Knudsen et al [58]</td>
</tr>
<tr>
<td>Interprofessional collaboration</td>
<td>2 (5)</td>
<td>Lam et al [41]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Moilanen et al [51]</td>
</tr>
<tr>
<td>Teamwork</td>
<td>7 (18)</td>
<td>Lam et al [41]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Al Mashmoum and Hamade [44]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Welter et al [47]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Patton [52]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fingrut et al [55]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mork-Knudsen et al [58]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thingnes and Lewis [59]</td>
</tr>
<tr>
<td>Category and facilitator</td>
<td>Studies, n (%)</td>
<td>Reference</td>
</tr>
<tr>
<td>------------------------------------------------------</td>
<td>----------------</td>
<td>---------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Multidisciplinary team</td>
<td>8 (21)</td>
<td>• Sharma et al [1]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Lam et al [41]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Kilsdonk et al [45]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Taba et al [43]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Rankin et al [60]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Kilsdonk et al [61]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Kane and Luz [62]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Kostaras et al [63]</td>
</tr>
<tr>
<td>Peer review</td>
<td>3 (8)</td>
<td>• Kilsdonk et al [45,61]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Mathews et al [64]</td>
</tr>
<tr>
<td>Web-based teaching</td>
<td>1 (3)</td>
<td>• Stoehr et al [65]</td>
</tr>
<tr>
<td>Web-based learning</td>
<td>2 (5)</td>
<td>• Armoogum and Buchgeister [48]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Shaw et al [66]</td>
</tr>
<tr>
<td>Learning</td>
<td>5 (13)</td>
<td>• Al Mashmoum and Hamade [44]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Welter et al [47]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Fatahi et al [50]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Thingnes and Lewis [59]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Obura et al [67]</td>
</tr>
<tr>
<td>Lectures, seminars, conferences, and journal club</td>
<td>2 (5)</td>
<td>• Adeyelure et al [36]</td>
</tr>
<tr>
<td>meetings</td>
<td></td>
<td>• Armoogum and Buchgeister [48]</td>
</tr>
<tr>
<td>Workshops and training</td>
<td>8 (21)</td>
<td>• Lisy et al [12]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Adeyelure et al [36]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Taba et al [42]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Bagayogo et al [49]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Fatahi et al [50]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Mork-Kundsen et al [58]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Samant et al [68]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Barbosa et al [69]</td>
</tr>
<tr>
<td>Extrinsic motivation</td>
<td>3 (8)</td>
<td>• Al Mashmoum and Hamade [44]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Kilsdonk et al [45]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Singh et al [46]</td>
</tr>
<tr>
<td>Physician rounds</td>
<td>2 (5)</td>
<td>• Adeyelure et al [36]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Fatahi et al [50]</td>
</tr>
</tbody>
</table>

**Technological facilitators**
<table>
<thead>
<tr>
<th>Category and facilitator</th>
<th>Studies, n (%)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>PACS(^a)</td>
<td>8 (21)</td>
<td>• Khajouei et al [18]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Adeyelure et al [36]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Taba et al [43]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Al Mashmoum and Hamade [44]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Welter et al [47]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Fatahi et al [50]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Thingnes and Lewis [59]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Stoehr et al [65]</td>
</tr>
<tr>
<td>ICT(^b)</td>
<td>3 (8)</td>
<td>• Adeyelure et al [36]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Taba et al [43]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Al Mashmoum and Hamade [44]</td>
</tr>
<tr>
<td>Network</td>
<td>6 (15)</td>
<td>• Taba et al [42,43]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Armoogum and Buchgeister [48]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Bagayogo et al [49]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Fingrut et al [55]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Addicott and Ferlie [70]</td>
</tr>
<tr>
<td>Social media</td>
<td>3 (8)</td>
<td>• Taba et al [43]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Singh et al [46]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Alanzi and Al-Habib [71]</td>
</tr>
<tr>
<td>Intranet and extranet</td>
<td>1 (3)</td>
<td>• Barbosa et al [69]</td>
</tr>
<tr>
<td>Multimedia and teleradiology</td>
<td>2 (5)</td>
<td>• Taba et al [43]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Al-Safadi [72]</td>
</tr>
<tr>
<td>Digital library</td>
<td>1 (3)</td>
<td>• Taba et al [43]</td>
</tr>
</tbody>
</table>

\(^a\) PACS: picture archiving and communication system.  
\(^b\) ICT: information and communications technology.
Table 2. Barriers affecting knowledge-sharing practices in medical imaging departments (n=39).

<table>
<thead>
<tr>
<th>Category and barrier</th>
<th>Studies, n (%)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Financial barriers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 (3)</td>
<td>Khajouei et al [18]</td>
</tr>
<tr>
<td><strong>Administrative barriers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Language</td>
<td>3 (8)</td>
<td>Adeyelure et al [36]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lam et al [41]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Stoehr et al [65]</td>
</tr>
<tr>
<td>Time</td>
<td>3 (8)</td>
<td>Adeyelure et al [36]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fatahi et al [50]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Glicksman et al [54]</td>
</tr>
<tr>
<td>Shortage of staff</td>
<td>1 (3)</td>
<td>Fatahi et al [50]</td>
</tr>
<tr>
<td>Less experience</td>
<td>1 (3)</td>
<td>Stoehr et al [65]</td>
</tr>
<tr>
<td>Lack of transparency</td>
<td>1 (3)</td>
<td>Lam et al [41]</td>
</tr>
<tr>
<td><strong>Technological barriers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Network</td>
<td>5 (13)</td>
<td>Khajouei et al [18]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Adeyelure et al [36]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Taba et al [42,43]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bagayogo et al [49]</td>
</tr>
<tr>
<td>Upgrade of system</td>
<td>1 (3)</td>
<td>Khajouei et al [18]</td>
</tr>
<tr>
<td>Lack of equipment</td>
<td>1 (3)</td>
<td>Khajouei et al [18]</td>
</tr>
<tr>
<td><strong>Geographical barriers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geographical distance</td>
<td>1 (3)</td>
<td>Armoogum and Buchgeister [48]</td>
</tr>
</tbody>
</table>

**Knowledge-Sharing Facilitators**

The identified knowledge sharing facilitators in medical imaging departments were classified into 3 categories, as previously mentioned: individual, departmental, and technological factors, shown in Table 1.

**Individual Facilitators**

Individual facilitators are considered the basic factors that allow knowledge-sharing practices to exist in any institution. These facilitators depend on the health care professionals’ attitudes in medical imaging departments. There were 10 facilitators identified that were related to individual factors.

The most cited facilitator was intrinsic motivation [44-48]. Kilsdonk et al [45] illustrated that the intrinsic motivation to participate is a key concept for external peer review programs to enhance knowledge sharing among health care professionals in multidisciplinary teamwork in cancer care. It was found in a cross-sectional survey that intrinsic motivation plays an essential role in knowledge sharing among health care professionals [46]. Moreover, intrinsic motivation depends on health care professionals’ needs and interests [47]. It was one of the important facilitators that enhanced knowledge sharing among health care professionals in medical imaging departments [44,48].

In total, 13% (5/39) of the studies illustrated that trust significantly affects knowledge sharing [43,44,49-51]. To facilitate knowledge sharing among interprofessional networks, trust building among them is needed [49,51]. Trust has not only a strong influence on knowledge sharing [43,44] but also a positive influence on communication between referring clinicians and radiologists [50]. Positive attitudes play a significant role in sharing knowledge among interprofessional collaborations [41]. Positive attitudes have been found to influence social networking among breast radiologists, which in turn influences knowledge sharing [42,43]. Moreover, Taba et al [42] reported that experience was a key characteristic of individual facilitators, which affects knowledge sharing. In addition, without awareness of the importance of knowledge sharing, it will not exist in medical imaging departments [44]. Taba et al [42,43] highlighted that self-esteem is considered one of the individual factors that have a positive impact on knowledge sharing. Self-efficacy and personality were reported in 5% (2/39) of the studies, which found that they play a significant role in sharing knowledge [44,46,52].

**Departmental Facilitators**

Departmental facilitators include resources, which are provided by medical imaging departments to enhance knowledge-sharing practices among their health care professionals. There were 21% (8/39) of the studies that concentrated on multidisciplinary teams (MDTs) [1,41,43,45,60-63]. Lam et al [41] illustrated
that interprofessional collaboration can improve knowledge sharing among them to increase patient outcomes. Interprofessional collaboration is defined as the process that occurs between multiple workers from different disciplines to achieve care for patients [41]. Moreover, Moilanen et al [51] showed that interprofessional collaboration plays a significant role in sharing knowledge and increasing well-being at the workplace. However, Lam et al [41] illustrated that MDTs were especially located in cancer centers to provide care for patients with cancer. These meetings were considered the best in cancer care and were very important for making decisions in Australia [1,60]. In the Netherlands, Kilsdonk et al [45,61] reported that regularly scheduled multidisciplinary meetings for sharing knowledge among medical professionals had a positive impact on making the best decisions regarding cancer cases. However, a lack of MDT meetings among health care professionals has a negative effect on patient outcomes in health care settings [53]. A survey that was conducted among MDT members showed that MDT meetings focused on sharing knowledge, collaborating, and making decisions among their specialized members. These members were from different disciplines, such as medical oncologists, radiologists, nurses, pathologists, physicians, coordinators, and radiation oncologists [62]. Radiologists reported that there were several benefits from MDT meetings, for example, gaining new knowledge and being able to discuss up-to-date information in the diagnosis of patients with cancer according to their disciplines [43]. In general, MDT meetings have positive effects on achieving consensus on diagnosis and treatment strategies based on knowledge sharing among their members [63].

There were 10% (4/39) of studies that reported that communities of practice (CoPs) have a direct impact on learning by enhancing knowledge-sharing behaviors among professional members [48,54-56]. Glicksman et al [54] highlighted that CoPs were increasingly used in the health care sector to improve patient outcomes by sharing knowledge among members. A total of 94% of interviewees reported that their experience in professional networks increased because of their involvement in the CoP [54]. Fingrut et al [55,56] showed in their study that tacit knowledge, which is the main type of knowledge, was shared during CoPs and that it is difficult to codify it. In addition, community of oncologists is used as a term to describe a CoP that plays a significant role in sharing information and knowledge among oncologists [53].

The importance of teamwork was reported in 18% (7/39) of the studies [41,44,47,52,55,58,59]. It has a significant role in knowledge sharing by allowing health care resources to be used in the proper way and minimizing service duplication [41]. Fingrut et al [55] reported that teamwork is very important to support collaboration with government services. Welter et al [47] illustrated that knowledge sharing takes place during teamwork to increase problem-solving strategies. According to the qualitative methods used in 5% (2/39) of the studies, interviews showed that teamwork can facilitate knowledge sharing among health care professionals in medical imaging departments [44,59]. To break the conflict among health care professionals in medical imaging departments, Patton [52] and Mork-Knudsen et al [58] reported that the role of the department is to enhance teamwork to manage workplace conflict by improving the departmental environment. Department arrangements have a positive impact on enhancing knowledge sharing by offering health care professionals the best office layout and an environment free of risk [44]. In addition, peer review is essential to improve teamwork in health care institutions and, therefore, knowledge sharing [45,61,64].

There were 21% (8/39) of studies that focused on the importance of training and workshops to support knowledge sharing [12,36,42,49,50,58,68,69]. In 2010, Armoogum and Buchgeister [48] illustrated in their survey, which took place in the radiology department, that 74% of respondents stated that seminars and journal clubs had a positive impact on supporting knowledge sharing. Adeyelure et al [36] reported that most South African health care centers play a significant role in encouraging their health care professionals to attend national and international conferences, workshops, and symposiums to facilitate knowledge sharing. Regularly attending workshops has a positive impact on knowledge sharing [50,68]. Taba et al [42] identified that breast radiology training has a positive impact on the work environment by facilitating knowledge sharing. Moreover, studies that reported training considered it a main means for knowledge-sharing accomplishment [12,69]. In addition, multidisciplinary training programs are crucial for facilitating knowledge sharing and interaction among professionals. In summary, the role of training focuses on achieving skills and maintaining a workplace environment, thereby facilitating knowledge-sharing practices among their health care professionals [58].

Web-based teaching in radiology departments played a significant role in enhancing knowledge sharing during the COVID-19 pandemic [65]. In addition, Adeyelure et al [36] and Fatahi et al [50] identified that physician rounds are considered another way of teaching and sharing knowledge among health care professionals, for example, physicians, nurses, and allied health professionals.

Learning plays a significant role in knowledge sharing, either attending physically or over the web. Web-based learning, or e-learning, enables collaborative knowledge sharing by using mobile devices or computers [66]. In 2010, Armoogum and Buchgeister [48] illustrated that web-based learning forms the shape of the body of knowledge sharing with radiology CoPs. In their studies, Welter et al [47], Al Mashmoun and Hamade [44], and Fatahi et al [50] reported that learning played a vital role in sharing tacit and explicit knowledge among health care professionals in radiology departments. The results of the survey conducted at a medical imaging department found that 95% of radiographers believed that learning and lifelong learning were important in radiography as they had a positive impact on sharing knowledge [59]. In general, learning occurs within a community to increase and support learning experiences by encouraging knowledge sharing among learners [67].

In the CoP, cultural collaboration plays a significant role in improving knowledge of outcomes of patients with cancer [55]. Taba et al [43] reported that departmental culture is very important among health care professionals in radiology departments as it has a strong impact on the workplace.
environment by enhancing frequently asking for further opinions. In their studies among radiographers, Barbosa et al [69] and Mork-Knudsen et al [58] showed that managers in a department have a huge responsibility to control the departmental culture by modifying it to create a strong environment for sharing health care professionals’ beliefs and thoughts and improving the practice of justification among them. This helps break the conflict among them as it encourages teamwork opportunities [52].

Leadership is the backbone of any department. Lee et al [57] reported that an empowering leader is crucial in decision-making and encourages health care professionals to share knowledge among themselves. In addition, leadership has a responsibility to support the department by allowing health care professionals to share their knowledge by building a healthy communication environment [35,44]. Furthermore, leaders play a crucial role in enhancing knowledge sharing and opening the door for creativity by breaking down conflict among health care professionals at the workplace [52]. In general, leaders have a huge responsibility to build a communication culture to enhance knowledge sharing [51].

Extrinsic motivation is a departmental facilitator that has a positive impact on knowledge sharing by providing rewards and reciprocal benefits [44-46].

**Technological Facilitators**

Technological facilitators include information and communications technologies (ICTs). The findings of 8% (3/39) of the studies indicated that ICT is considered a core facilitator for building professional social networks and enhancing work environment practices among health care professionals in medical imaging departments [36,43,44]. In total, 21% (8/39) of the studies identified the role of the PACS in knowledge-sharing practices in medical imaging departments [18,36,43,44,47,50,59,65]. Stoehr et al [65] showed that the PACS played a vital role in web-based conferences by making cases and tumors obvious in an easy way. The findings of a qualitative study revealed that the PACS has advantages in the transmission, reception, retrieval, processing, distribution, and display of medical reports and imaging from one workstation to another [43].

Several studies (5/39, 13%) found that interprofessional networks are important for facilitating knowledge sharing and improving patient outcomes [49,55]. This network could be either an intranet within a department or an extranet between one department and another [69]. In 2007, Addcott and Ferlie [70] demonstrated that networks were considered means of sharing knowledge and good practice among professionals from different disciplines and other health care institutions involved in patient care. Furthermore, a network exists in all radiology departments as it is a strong knowledge-sharing practice and positively affects the workplace [43,48]. Teleradiology and internet-based multimedia interaction play a vital role in knowledge sharing by transmitting radiographic imaging and written or spoken words from one location to another, for example, multimedia internet-based and teleradiology management systems [43,72].

Social media platforms are considered a useful tool for health care professionals [43,46]. In the results of a survey study among health care professionals, >80% of respondents stated the importance of social media in improving knowledge sharing, thereby improving decision-making [71].

The establishment of digital libraries has changed the way the radiology environment operates and the way of searching for information [43]. Interviews reported that electronic resources such as e-books and databases could support making decisions. Moreover, it is an effective source for education and solving work-related problems [43].

This section identified the facilitators that affect knowledge sharing among health care professionals in medical imaging departments. The section that follows will identify the barriers that hinder knowledge-sharing practices.

**Knowledge-Sharing Barriers**

These barriers are shown in Table 2.

Financial barriers are considered one of the main barriers that have a negative impact on knowledge-sharing behaviors, for example, the low cost to support ICT tools [18]. In addition to financial barriers, there were administrative barriers that hindered knowledge-sharing practices. Language barriers were found to be a source of reluctance on knowledge sharing among health care professionals in medical imaging departments in 8% (3/39) of the studies [36,41,65]. In addition, health care professionals were reluctant to share knowledge as they did not have enough experience to share it with others [65]. Lam et al [41] reported in their study that a lack of transparency could affect knowledge sharing because of a lack of awareness of departmental policies and visions. Furthermore, there was reduced knowledge sharing among health care professionals because of a shortage of staff [50]. Moreover, time constraints were considered a barrier that impeded knowledge sharing as health care professionals who had many tasks to achieve often did not have enough time to share knowledge [36,50,54].

Technology plays a significant role in aiding the knowledge-sharing process. However, in several studies (5/39, 13%), it was found that low-speed networks had a negative impact on knowledge sharing as most tools that support knowledge sharing require a high-speed network [18,36,42,43,49]. Khajouei et al [18] showed that the lack of equipment and support for upgrading systems affected knowledge-sharing behaviors. Finally, the distance between geographically spread health care professionals caused communication issues [48].

**Discussion**

**Principal Findings**

**Overview**

This study identified the factors that affect knowledge sharing in medical imaging departments in cancer centers. The analysis of the selected 39 articles revealed that medical imaging departments have several facilitators and barriers affecting the knowledge-sharing process. All those facilitators and barriers
can apply to all medical imaging departments in general hospitals and cancer centers. All the selected studies (39/39, 100%) were conducted in medical imaging departments. However, 67% (26/39) were conducted in cancer centers.

**Knowledge-Sharing Facilitators in Medical Imaging Departments in General Hospitals Versus Cancer Centers**

The findings of this study revealed that all facilitators can apply to all medical imaging departments in general hospitals and cancer centers. However, some of the terminology of facilitators is different in medical imaging departments in cancer centers because of the nature of dealing with cancer cases in these centers. The differences in facilitator terminology between medical imaging departments in general hospitals and cancer centers are shown in Table 3.

<table>
<thead>
<tr>
<th>Type of facilitator</th>
<th>Facilitator terminology in medical imaging departments in general hospitals</th>
<th>Facilitator terminology in medical imaging departments in cancer centers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual facilitators</td>
<td>Trust  • Positive attitudes  • Awareness  • Experience  • Intrinsic motivation  • Personality  • Self-esteem  • Self-efficacy</td>
<td>Trust  • Positive attitudes  • Awareness  • Experience  • Intrinsic motivation  • Personality  • Self-esteem  • Self-efficacy</td>
</tr>
<tr>
<td>Departmental facilitators</td>
<td>CoPs(^a) and interprofessional collaboration  • Leadership  • Culture  • Teamwork  • Extrinsic motivation  • Learning and training  • Physician rounds  • Departmental arrangements</td>
<td>MDT(^b) and community of oncologists  • Leadership  • Culture  • Teamwork  • Extrinsic motivation  • Learning and training  • Physician rounds  • Departmental arrangements</td>
</tr>
<tr>
<td>Technological facilitators</td>
<td>ICT(^c) (PACS(^d), social media, intranet, extranet, telemedicine, and teleradiology)  • Network  • Digital library</td>
<td>• ICT (PACS, social media, intranet, extranet, telemedicine, and teleradiology)  • Network  • Digital library</td>
</tr>
</tbody>
</table>

\(^a\)CoP: community of practice.  
\(^b\)MDT: multidisciplinary team.  
\(^c\)ICT: information and communications technology.  
\(^d\)PACS: picture archiving and communication system.

First, individual facilitators play a significant role in enhancing knowledge sharing in medical imaging departments. They comprise several facilitators that are consistent in medical imaging departments be it in a cancer center or not. Trust has been proven to be an important determinant of knowledge sharing. Trust is the backbone of any relationship, so it enables knowledge sharing in medical imaging departments, especially tacit knowledge. Building trust among health care professionals who have less experience is important to enhance knowledge sharing among those who are experts in their field [49]. Fatahi et al [50] reported several actions to increase interprofessional trust, for example, face-to-face communication and phone contacts between referring clinicians and radiologists.

The importance of intrinsic motivation, which is related to knowledge sharing, could be observed in several studies (5/39, 13%). Intrinsic motivation has a direct impact on knowledge-sharing attitudes [46]. For instance, when health care professionals are not motivated to share what they have, they tend to keep the knowledge to themselves. In addition, positive attitudes are directly related to existing knowledge-sharing behaviors as they motivate health care professionals to share knowledge. The awareness of the importance of knowledge sharing among health care professionals is important to encourage them to share knowledge frequently to increase patient outcomes. Furthermore, personality is considered an individual facilitator that enhances knowledge sharing such that health care professionals who have positive attitudes tend to share knowledge with their peers. Self-efficacy and self-esteem are also important traits that motivate health care professionals to share their knowledge. In general, individual facilitators are crucial in medical imaging departments to build knowledge-sharing environments.

Second, there are departmental facilitators that enhance knowledge sharing among health care professionals in medical imaging departments. The existence of these facilitators is directly related to the success of the departmental policies. Although these facilitators are the same in all medical imaging departments, the terminology that describes CoPs in cancer
centers is different. These are called MDTs and community of oncologists.

Culture has been identified as a vital departmental facilitator that enables knowledge-sharing practices. In addition, Fingrut et al [55] reported that cultural communication plays a very important role in building a CoP to improve cancer care. Culture is a powerful facilitator to share knowledge by creating a healthy environment for innovation, community, and freedom to ask questions [43]. Leadership plays a crucial role in enhancing knowledge sharing among health care professionals in medical imaging departments, for example, breaking down conflict among them by understanding them and giving them opportunities to work with each other in a healthy environment [52]. Moreover, a leader is responsible for building their trust and motivating them to share knowledge among them. The concept of empowering leadership in relation to knowledge sharing was observed in the study by Lee et al [57]. This study illustrated that empowering leadership plays a vital role in promoting knowledge-sharing behaviors among health care professionals. Administration arrangements tend to affect the transfer of both types of knowledge by offering health care professionals spaces and offices to share their knowledge in a proper way [44]. Furthermore, both intrinsic and extrinsic motivation have a positive impact on knowledge-sharing attitudes by giving health care professionals awards and bonuses for sharing knowledge [44,46].

From the analytical review of the articles, there are several communities that support knowledge sharing in medical imaging departments. In general hospitals, the most popular community is called the CoP. It has become more common throughout medical imaging departments to share knowledge, with the major goal of enhancing the quality of services [54]. In addition, interprofessional collaboration is another type of community that enhances knowledge sharing through collaboration among several professionals with different knowledge backgrounds to achieve the highest quality of care [41]. However, in cancer centers, there are several communities, for instance, the community of oncologists. Dicicco-Bloom and Cunningham [53] illustrated that the purpose of this community is to give oncologists the chance to share their knowledge regarding special cancer care to improve patient outcomes. Furthermore, there is the MDT meeting. This kind of meeting plays a significant role in sharing knowledge among professionals to make a proper decision regarding specific cancer cases that are involved in this meeting based on several requirements to select the patient in question [33]. In general, teamwork, either within a community or in a separate group, plays an obvious role in building strong knowledge-sharing behaviors in radiology departments [58].

Workshops and training such as lectures, seminars, conferences, and journal club meetings are essential to circulate tacit and explicit knowledge among health care professionals in medical imaging departments [48]. Medical imaging departments should organize these activities annually (such as conferences), monthly (such as workshops), weekly (such as lectures and journal club meetings), and daily (such as morning sessions) to create an active environment for sharing knowledge among health care professionals. Furthermore, learning in radiology centers plays an essential role in making decisions by developing radiologists’ ability to use the available tools (eg, PACS) to retrieve images to share with other colleagues [47]. During the COVID-19 pandemic, web-based learning became predominant because of social distancing. Therefore, web-based learning is the best tool to enhance knowledge sharing among health care professionals without having to consider geographical barriers.

Finally, there are technological facilitators that affect knowledge-sharing practices in medical imaging departments. These facilitators are consistent in medical imaging departments whether in general hospitals or cancer centers. They have a positive impact on knowledge sharing in medical imaging departments. The role of ICT in knowledge sharing has become very important because of the teleological revolution. The most cited type of ICT facilitator was the PACS, which is well-known in medical imaging departments. The PACS is a powerful tool that encourages knowledge sharing among health care professionals by providing them with the ability to send and receive many reports and images of different patient cases from one location to another [18]. This interaction to share knowledge can happen within a department or among different departments [36]. This type of facilitator is used only in medical imaging departments. There are 2 ways to facilitate internet-based intranets or extranets [69]. Although technological facilitators are important, high-speed networks are required to perform several tasks in a proper way. For instance, the UK National Health Service has established Managed Clinical Networks especially for cancer cases to streamline patient pathways and increase knowledge sharing among professionals who are involved in cancer care [70]. Social media is another example of a technological facilitator that is part of ICTs and enhances and facilitates formal and informal knowledge sharing in health care institutions. Social media such as Facebook, WhatsApp, blogs, and wikis has a strong impact on enhancing knowledge sharing among health care professionals in the health sector. The results of the survey by Alanzi and Al-Habib [71] showed that social media is a powerful instrument to enhance teaching that has a positive role in making decisions and solving problems. In addition, telemedicine and teleradiology play a significant role in enhancing knowledge sharing among health care professionals by sharing images of the scans among them to interpret an appropriate report [43,72]. Digital libraries are instrumental in enhancing knowledge sharing as they play a vital role in learning and problem-solving [43].

Knowledge-Sharing Barriers in Medical Imaging Departments in General Hospitals Versus Cancer Centers

In addition to the facilitators that affect knowledge sharing, there are several barriers that hinder knowledge-sharing practices. These barriers can apply to all medical imaging departments in general hospitals and cancer centers, as shown inTextbox 2. Financial barriers such as costs are considered one of the most predominant barriers that affect knowledge sharing [18]. The PACS, hospital information systems, and registration information systems require a large amount of money for upgrading and maintenance to work efficiently without losing patient information [18].
Textbox 2. Barriers that hinder knowledge sharing in medical imaging departments in general hospitals and cancer centers.

<table>
<thead>
<tr>
<th>Financial barriers</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Cost</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Administrative barriers</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Language</td>
</tr>
<tr>
<td>• Time</td>
</tr>
<tr>
<td>• Shortage of staff</td>
</tr>
<tr>
<td>• Lack of transparency</td>
</tr>
<tr>
<td>• Lack of experience</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Technological barriers</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Low speed in network</td>
</tr>
<tr>
<td>• Upgrade of the system</td>
</tr>
<tr>
<td>• Lack of equipment</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Geographical barriers</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Geographical distance</td>
</tr>
</tbody>
</table>

The studies showed that there are administrative barriers that have a negative impact on knowledge-sharing behaviors, such as language barriers, time constraints, lack of experience, shortage of staff, and lack of transparency. The language barrier is the main barrier facing administration as language is the first route for health care professionals to communicate with their peers and share their knowledge [65]. Therefore, the administration should select a language that suits the majority. In addition, time constraints are a barrier that hinders knowledge sharing [54]. Insufficient time did not allow health care professionals in medical imaging departments to share their knowledge as they were busy with cases all the time. Thus, the administration should offer them free time to share their knowledge by attending meetings. Lam et al [41] illustrated that a lack of transparency impedes knowledge-sharing practices as the administration does not have a clear policy or framework to activate knowledge-sharing behaviors. Moreover, experts have a tendency to share their knowledge more than those who have less experience [65]. To avoid that, conducting educational practices is vital to encourage health care professionals to gain new experiences and keep them up-to-date.

There were several technological barriers, but the most cited one was networks. Poor networks can hinder not only knowledge-sharing practices but also health care procedures [18,36]. In addition, the lack of equipment has a negative impact on knowledge sharing. Maintaining and upgrading systems is essential to enhance knowledge sharing among health care professionals in medical imaging departments [18].

The distance between geographically separated health care professionals worldwide acts as a barrier and causes communication problems [48]. Knowledge sharing among health care professionals becomes easier when they meet without a geographical barrier or if the physical distance is not a concern. However, with the growth of web-based meetings, especially during the COVID-19 pandemic, teaching, learning, and meeting over the web are useful tools to maintain knowledge-sharing practices among health care professionals and break down these barriers.

This study identified factors that affect knowledge sharing in medical imaging departments in cancer centers and general hospitals. All facilitators and barriers can apply to medical imaging departments in general hospitals and cancer centers. However, we note that the terminology used to describe facilitators of and barriers to knowledge sharing is inconsistent across health care sectors depending on the facilitators and the nature of the work in those sectors. For example, in medical imaging departments in cancer centers, MDT meetings and communities of oncologists are considered a type of CoP. They constitute departmental facilitators and are used frequently in cancer centers [1,41,43,45,53,60-63].

The findings of this review are consistent with those of other studies on the factors that affect knowledge sharing in all health care settings [19-32]. This is presumably because this study focused on health care sectors, which have the same environment. This environment has demonstrated the interaction of tacit and explicit knowledge among health care professionals to share knowledge that depends on several factors. Although these factors have remained consistent, the PACS is only used in medical imaging departments, but the remaining factors can apply to different departments in general hospitals.

Limitations and Strengths

There were several limitations to this study that should be acknowledged. There were 7 search engines that were used in this systematic review. Although these databases are relevant for health care publications, there is a possibility that unrelated studies were included. In addition, a few databases had a small number of results. As this study was restricted to only medical imaging departments, we could not determine whether the factors that affect knowledge sharing in those departments are
the same for all departments in health care settings. Further work is required to assess this. To the best of our knowledge, there has been no previous systematic review that identified factors affecting knowledge-sharing practices among health care professionals in medical imaging departments in cancer centers.

Conclusions
This systematic review revealed the factors that can serve as a framework for facilitating the overall knowledge-sharing process in any medical imaging department in a general hospital or cancer center. In terms of the facilitators of and barriers to knowledge sharing, this study showed that they are the same in medical imaging departments, whether in cancer centers or general hospitals. However, the terminologies might be different based on the nature of these departments. Medical imaging departments exist as part of health care services, and they have several tasks that have increased gradually because of advances in technology and imaging procedures, for instance, implementing new technologies for imaging and diagnosing patients’ conditions.

This study identified a source of knowledge for medical imaging departments and a clear understanding of facilitators and barriers that affect knowledge-sharing practices. Therefore, the managers and policy makers of medical imaging departments should be aware of these facilitators and barriers to create a framework that enhances knowledge sharing and avoids any challenges health care professionals might face regarding the knowledge-sharing process. Furthermore, it will inform them of the deficiencies in knowledge management implementation because of the lack of an effective knowledge-sharing process.

Acknowledgments
The Medical Subject Heading terms with the Boolean operators were developed with the help of a Teaching and Learning Librarian from the University of Manchester, Micheal Stevenson. This study was completed as part of doctoral studies funded by the Ministry of Health, Kuwait, and the University of Manchester, United Kingdom.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Results of the search strategies used in PubMed Central, EBSCOhost (CINAHL), Ovid MEDLINE, Ovid Embase, Elsevier (Scopus), ProQuest, and Clarivate (Web of Science).

[DOCX File, 26 KB - humanactors_v10i1e44327_app1.docx ]

Multimedia Appendix 2
Characteristics of the included studies.

[DOCX File, 37 KB - humanactors_v10i1e44327_app2.docx ]

Multimedia Appendix 3

[DOCX File, 54 KB - humanactors_v10i1e44327_app3.docx ]

References


25. 10.1186/1472-6963-14-126


71. Almashmoum et al. JMIR HUMAN FACTORS 2023 | vol. 10 | e44327 | p.301https://humanfactors.jmir.org/2023/1/e44327 (page number not for citation purposes)

Abbreviations

- CoP: community of practice
- ICT: information and communications technology
- MDT: multidisciplinary team
- MMAT: Mixed Methods Appraisal Tool
- PACS: picture archiving and communication system
- PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
Evaluating the Usability of an Emergency Department After Visit Summary: Staged Heuristic Evaluation

Hanna J Barton\textsuperscript{1,2}, BSc, MEng; Megan E Salwei\textsuperscript{3,4}, PhD; Rachel A Rutkowski\textsuperscript{1,2}, PhD; Kathryn Wust\textsuperscript{1,2}, BSc, MEng; Sheryl Krause\textsuperscript{5}, MSc, RN, CEN, ACNS-BC; Peter LT Hoonakker\textsuperscript{1}, PhD; Paula vW Dail\textsuperscript{6}, PhD; Denise M Buckley\textsuperscript{7}, BSN; Alexis Eastman\textsuperscript{8,9}, MD; Brad Ehlenfeldt\textsuperscript{7}, BBA; Brian W Patterson\textsuperscript{10}, MD; Manish N Shah\textsuperscript{10}, MD; Barbara J King\textsuperscript{5}, PhD; Nicole E Werner\textsuperscript{11}, PhD; Pascale Carayon\textsuperscript{1,2}, PhD

\textsuperscript{1}Wisconsin Institute for Healthcare Systems Engineering, University of Wisconsin-Madison, Madison, WI, United States
\textsuperscript{2}Department of Industrial and Systems Engineering, University of Wisconsin-Madison, Madison, WI, United States
\textsuperscript{3}Center for Research and Innovation in Systems Safety, Department of Anesthesiology, Vanderbilt University Medical Center, Nashville, TN, United States
\textsuperscript{4}Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, United States
\textsuperscript{5}School of Nursing, University of Wisconsin-Madison, Madison, WI, United States
\textsuperscript{6}University of Wisconsin-Madison Health Sciences Patient and Family Advisory Council Member, Madison, WI, United States
\textsuperscript{7}Berbee Walsh Department of Emergency Medicine, University of Wisconsin Hospital and Clinics, Madison, WI, United States
\textsuperscript{8}Center for Aging Research and Education, School of Nursing, University of Wisconsin-Madison, Madison, WI, United States
\textsuperscript{9}Department of Medicine, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, WI, United States
\textsuperscript{10}Berbee Walsh Department of Emergency Medicine, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, WI, United States
\textsuperscript{11}Department of Health and Wellness Design, Indiana University School of Public Health-Bloomington, Bloomington, IN, United States

Corresponding Author:
Hanna J Barton, BSc, MEng
Wisconsin Institute for Healthcare Systems Engineering
University of Wisconsin-Madison
3139 Engineering Centers Building
1550 Engineering Drive
Madison, WI, 53706
United States
Phone: 1 6083586120
Email: hbarton@wisc.edu

Abstract

Background: Heuristic evaluations, while commonly used, may inadequately capture the severity of identified usability issues. In the domain of health care, usability issues can pose different levels of risk to patients. Incorporating diverse expertise (eg, clinical and patient) in the heuristic evaluation process can help assess and address potential negative impacts on patient safety that may otherwise go unnoticed. One document that should be highly usable for patients—with the potential to prevent adverse outcomes—is the after visit summary (AVS). The AVS is the document given to a patient upon discharge from the emergency department (ED), which contains instructions on how to manage symptoms, medications, and follow-up care.

Objective: This study aims to assess a multistage method for integrating diverse expertise (ie, clinical, an older adult care partner, and health IT) with human factors engineering (HFE) expertise in the usability evaluation of the patient-facing ED AVS.

Methods: We conducted a three-staged heuristic evaluation of an ED AVS using heuristics developed for use in evaluating patient-facing documentation. In stage 1, HFE experts reviewed the AVS to identify usability issues. In stage 2, 6 experts of varying expertise (ie, emergency medicine physicians, ED nurses, geriatricians, transitional care nurses, and an older adult care partner) rated each previously identified usability issue on its potential impact on patient comprehension and patient safety. Finally, in stage 3, an IT expert reviewed each usability issue to identify the likelihood of successfully addressing the issue.

Results: In stage 1, we identified 60 usability issues that violated a total of 108 heuristics. In stage 2, 18 additional usability issues that violated 27 heuristics were identified by the study experts. Impact ratings ranged from all experts rating the issue as “no impact” to 5 out of 6 experts rating the issue as having a “large negative impact.” On average, the older adult care partner...
representative rated usability issues as being more significant more of the time. In stage 3, 31 usability issues were rated by an IT professional as “impossible to address,” 21 as “maybe,” and 24 as “can be addressed.”

Conclusions: Integrating diverse expertise when evaluating usability is important when patient safety is at stake. The non-HFE experts, included in stage 2 of our evaluation, identified 23% (18/78) of all the usability issues and, depending on their expertise, rated those issues as having differing impacts on patient comprehension and safety. Our findings suggest that, to conduct a comprehensive heuristic evaluation, expertise from all the contexts in which the AVS is used must be considered. Combining those findings with ratings from an IT expert, usability issues can be strategically addressed through redesign. Thus, a 3-staged heuristic evaluation method offers a framework for integrating context-specific expertise efficiently, while providing practical insights to guide human-centered design.

(JMIR Hum Factors 2023;10:e43729) doi:10.2196/43729

KEYWORDS

patient safety; heuristic evaluation; usability; emergency medicine; safety; emergency; human factors engineering; usability; discharge summary; documentation; heuristic

Introduction

Overview

Heuristic evaluations are commonly used to evaluate the usability of health technologies [1,2]. Relying on human factors or usability experts to assess a technology against usability criteria (ie, heuristics), heuristic evaluations offer an efficient and cost-effective alternative to user-based evaluation methods [3]. However, the method’s reliance on human factors expertise may limit its applicability and usefulness, especially regarding the evaluation of the severity of identified usability violations. In the domain of health care, usability violations can pose different levels of risk or harm to the patient; therefore, heuristic evaluation may require additional expertise besides human factors expertise [4,5]. One solution to this challenge is integrating other domains of expertise, such as clinical, patient and care partner, and IT expertise in the evaluation of a technology’s usability.

Background

Selection of a list of criteria—whether referred to as guidelines, design principles, or heuristics—that constitute a “usable” technology is an essential aspect of conducting a heuristic evaluation. Molich and Nielsen’s [2] 1990 seminal article introducing heuristic evaluation included initial principles: simple and natural dialogue, speak the user’s language, minimize the user’s memory load, be consistent, provide feedback, clearly marked exits, shortcuts, good error messages, and error prevention. In practice, Nielsen’s [6] 10 heuristics, published online in 1995, are the most frequently used.

Typically, in conducting a heuristic evaluation, 1 expert reviews the technology looking for any and all violations of the selected usability criteria, producing a list of usability violations. Some identified violations are less significant than others, and as such, a follow-up step is often used to assess the severity of each violation to give direction for prioritization and redesign efforts. Upon initial conceptualization by Nielsen [7], a 5-step severity scale is often applied with scores that range from 0 (“not a usability problem at all”) to 4 (“usability catastrophe”).

Adapting Heuristic Evaluation

Heuristic evaluations have been adapted for many domains and technologies, typically in one of the following ways: (1) the usability criteria on which the technology is evaluated, (2) the evaluation of the severity of usability violations, and (3) the mode of conducting the evaluation (eg, in groups) [3,5,8].

For example, Zhang et al [5] adapted the heuristic evaluation method for the assessment of medical devices, developing the Nielsen-Schneiderman heuristics. A synthesis of Nielsen’s 10 heuristics with Schneiderman’s “eight golden rules,” the 14 Nielsen-Schneiderman heuristics and their subbullets provide a comprehensive list of usability criteria that are especially useful in the evaluation of medical devices and health IT [5]. Another variation of the usability criteria is the ergonomic criteria defined by Scapin and Bastien [3]. They outline 8 categories of usability criteria: guidance, workload, explicit control, adaptability, error management, consistency, significance of codes, and compatibility. In contrast to the Nielsen-Schneiderman heuristics, the ergonomic criteria of Scapin and Bastien [3] provide a broader, macro-view of usability including consideration of workflow integration seen by their criteria “compatibility.”

Hermawati and Lawson [9] distinguish between general heuristics and heuristics developed for specific domains such as the evaluation of the usability of patient-facing documentation. For example, Tremoulet et al [8] conducted a heuristic evaluation of an emergency department (ED) after visit summary (AVS), the document handed to patients as they are discharged from the ED, that contains instructions and information to help them manage their symptoms, medications, and follow-up care [10]. Aiming to evaluate the usability of the AVS by outpatient clinicians (eg, clinicians supporting follow-up care), the authors adapted heuristic evaluation in a few ways. First, they selected usability criteria that integrated Nielsen’s heuristics with guidelines for effective health communication, so that the usability of the document could be more accurately assessed [8]. Further, consistent with participatory ergonomics principles, they partnered with clinical and human factors experts to conduct the evaluations [11]. In total, they identified 224 distinct usability issues across the 4 AVS documents they reviewed, of which 12 were considered catastrophic. For each of the AVS reviewed, content issues (eg,
clarity of content, emphasis, context, relevance, and absence or lack of information) were the most common, accounting for 32% of the identified violations.

While Tremoulet et al [8] offer a domain-specific list of heuristics (ie, for the patient-facing AVS) and a compelling method for including clinical experts (eg, primary care physicians) in the heuristic evaluation of patient-facing documents, there remains a gap in understanding the usability of the AVS from the patient’s perspective. This is important as the patient is the one who will ultimately receive the AVS (presumably), carry it home, and (possibly) interact with it after discharge from the ED. Further, the AVS has been identified as an important tool for care coordination between the ED and the home—a transition that is highly vulnerable to patient safety problems [10,12,13].

Therefore, in evaluating the usability of patient-facing documents, it is important to include the perspective of patients and care partners, as well as the perspective of clinicians who interact with patients and care partners in sharing and reviewing those documents. In addition, the heuristic evaluation can produce more impactful results if the violations are evaluated for potential redesign; this calls for the involvement of IT experts who can provide important information on whether violations can be addressed in the redesign phase. Thus, adapting heuristic evaluation methods to efficiently incorporate expertise from clinicians, patients and care partners, and IT professionals is necessary to assess and address potential impacts on patient safety.

Research Objective
The objective of this study was to assess a method for integrating diverse expertise (ie, clinical, patient and care partner, and IT) with human factors engineering (HFE) expertise in the evaluation of an ED AVS.

Methods
Overview
This study was part of an AHRQ Patient Safety Learning Lab aimed at developing a set of tools to improve care coordination for older adults who come to the ED with a fall or suspected urinary tract infection [14]. As part of the development of an intervention to improve the discharge process for patients transitioning to the home, we recognized the need for an initial assessment of the patient-facing ED AVS. As such, we conducted a 3-staged heuristic evaluation (Table 1) of 2 versions of an ED AVS to inform the design and implementation of a patient-centered discharge process. This work was done early in the COVID-19 pandemic, and therefore, was conducted digitally via videoconferencing software.

Table 1. Three-staged heuristic evaluation method.

<table>
<thead>
<tr>
<th>Stage</th>
<th>Expertise</th>
<th>Guiding questions</th>
<th>Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>HFE(^a)</td>
<td>What usability issues exist in the current AVS(^b)? What heuristics do they violate?</td>
<td>1.5-hour meeting with 8 HFE experts facilitated by 1 researcher (HJB)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sent preparation materials to HFE experts: key literature, the AVS documents being evaluated, list of heuristics, and an example output of a heuristic evaluation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Using the AVS documents provided and moving from left to right and from top to bottom, the group identified usability issues and the specific heuristics they violated</td>
</tr>
<tr>
<td>2</td>
<td>Clinical (emergency medicine, geriatrics, and nursing) patient and care partner</td>
<td>What issues have the largest impact on patient safety and comprehension? What do we need to address first?</td>
<td>Six participants: 2 emergency medicine physicians, 1 ED(^c) nurse, 1 nurse with transitional care expertise, 1 geriatrician, and 1 older adult care partner</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Participants rated each usability issue as having “no impact,” “some negative impact,” or “large negative impact” on our 2 criteria: patient comprehension and patient safety (~1 hour)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>30-minute one-on-one debrief with each participant led by HFE team members (HJB and PC) to resolve outstanding questions and capture additional usability issues</td>
</tr>
<tr>
<td>3</td>
<td>Health IT</td>
<td>What issues can we address?</td>
<td>IT expert scored each violation as “can be addressed,” “maybe,” or “impossible to address” in response to the prompt: “How likely are we (from an IT perspective) to be able to address this violation?” (~1 hour)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>30-minute one-on-one debrief with participants led by HFE team members (HJB and PC) to ask clarifying questions</td>
</tr>
</tbody>
</table>

\(^a\)HFE: human factors engineering.

\(^b\)AVS: after visit summary.

\(^c\)ED: emergency department.

https://humanfactors.jmir.org/2023/1/e43729
Selection of Heuristics

Heuristics for evaluating the ED AVS were selected by comparing Tremoulet et al’s [8] domain-specific heuristics to 2 prominent sets of heuristics, discussed in the introduction: (1) Scapin and Bastien’s [3] list of ergonomic criteria and (2) the Nielsen-Schneiderman heuristics developed by Zhang et al [5]. The results of this comparison demonstrated that the Tremoulet et al [8] heuristics were comprehensive, and yet tailored for the evaluation of paper-based, patient-facing documentation. Thus, we selected the Tremoulet et al [8] heuristics, modifying them slightly to include questions from the associated Scapin and Bastien [3] and Nielsen-Schneiderman heuristics (Table 2).


<table>
<thead>
<tr>
<th>Heuristic categories and names</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Readability</td>
<td>The information is presented in a manner that is easy to read.</td>
</tr>
<tr>
<td>Color and contrast</td>
<td>Does the text have sufficient contrast?</td>
</tr>
<tr>
<td>Layout and position</td>
<td>Is the layout appealing, clear, and consistent across the document?</td>
</tr>
<tr>
<td>Font and capitalization</td>
<td>Are the font and its size consistent and readable?</td>
</tr>
<tr>
<td>Structure and format</td>
<td>Are the structure and format of each section effective and uniform?</td>
</tr>
<tr>
<td>Minimalism</td>
<td>Information is presented as simply and succinctly as possible.</td>
</tr>
<tr>
<td>Simple and direct</td>
<td>Are the language and sentence structure simple, direct, specific, and concise? <em>Note: Simple is not equivalent to abstract and general</em></td>
</tr>
<tr>
<td>Progressive level of detail</td>
<td>Does the document present the most important information first, following with increasing levels of detail?</td>
</tr>
<tr>
<td>Comprehensibility</td>
<td>It is easy for the reader to make sense of the information that is presented.</td>
</tr>
<tr>
<td>Terminology</td>
<td>Are complex and technical terms used correctly and consistently? <em>Are standard meanings of words used? Is language from the users’ perspective?</em></td>
</tr>
<tr>
<td>Clarity of headings</td>
<td>Are the headings clear and understandable?</td>
</tr>
<tr>
<td>Content</td>
<td>All the information that is presented is relevant to either a clinical expert or the older adult care partner, and no information needed by either of these parties is missing.</td>
</tr>
<tr>
<td>Clarity of content</td>
<td>Is the purpose of the material obvious?</td>
</tr>
<tr>
<td>Emphasis</td>
<td>Are important points emphasized appropriately? Is it clear why certain text is emphasized?</td>
</tr>
<tr>
<td>Context</td>
<td>Does the document include the creation or printing date and contact information?</td>
</tr>
<tr>
<td>Relevance</td>
<td>Is the content relevant to the patient’s condition and context? Is there extraneous information?</td>
</tr>
<tr>
<td>Absence or lack of information</td>
<td>Is any important content missing?</td>
</tr>
<tr>
<td>Organization</td>
<td>Information is ordered logically and grouped into reasonably sized sections with prominent and meaningful headings and subheadings.</td>
</tr>
<tr>
<td>Grouping</td>
<td>Is the information grouped in a meaningful format? Are the groups reasonably sized? <em>Is there clear visual distinction between sections?</em></td>
</tr>
<tr>
<td>Order</td>
<td>Is the information ordered logically? <em>Is like content grouped together?</em></td>
</tr>
<tr>
<td>Use of subheadings</td>
<td>Does the document use prominent and meaningful headings and subheadings?</td>
</tr>
<tr>
<td>Navigational tools</td>
<td>Does the material have navigational tools to help orient the reader? <em>Is context-sensitive help embedded in the contents?</em></td>
</tr>
</tbody>
</table>

Selection of ED AVS

For our heuristic evaluation, an ED AVS was simulated with fake patient data. In addition, we evaluated a redacted real-life ED AVS provided by the care partner who participated in our study. Using the second ED AVS allowed us to identify any usability issues that were artifacts of the simulation.

Stage 1: HFE Experts Identify Usability Issues

To identify usability issues, a group of 8 HFE experts met for 1.5 hours on June 23, 2020, to review the 2 AVS. Before conducting the evaluation, all participants were asked to review Tremoulet et al [8] article, the finalized list of heuristics (Table 2), the AVS documents being evaluated, and an example of a final report produced from a past heuristic evaluation. Additional heuristics literature was provided for the participants to review if they elected to [2,5].

During the virtual meeting, 1 researcher served as a facilitator (HJB)—sharing their screen and guiding the group through the ED AVS document from top-to-bottom and left-to-right. All participants were encouraged to verbalize the usability issues they viewed. Once an issue was identified, participants worked collaboratively to name the associated heuristics violated. When issues were identified, the facilitator circled them on the shared view of the ED AVS and numbered them for ease of reference.
The final list of identified usability issues and their associated heuristic violations was reconciled by researchers (HB, KW, and RR) within 24 hours of the group meeting. Snipped images of the marked-up ED discharge summaries were taken to give context for each of the issues identified.

**Stage 2: Clinical, Patient, and Care Partner Experts Rate the Impact of Usability Issues**

We selected a variety of experts to assess the impact of the identified usability issues on patient comprehension and safety. These experts include emergency medicine physicians (n=2), an ED nurse (n=1), a nurse with transitional care expertise (n=1), a primary care geriatrician (n=1), and an older adult care partner (n=1).

The type of expertise each expert provided was unique. The care partner referred to their perspective as an older adult and their lived experience having previously visited the ED with their partner 14 times over the course of 10 weeks. The emergency medicine physicians and ED nurse used their clinical expertise; the ED nurse also referred to nurses’ experiences reviewing the AVS with patients and their care partners as they are being discharged from the ED. Further, a nurse with expertise in older adult transitions and a geriatrician provided perspective on how patients and their care partners interact (or do not interact) with the AVS after discharge from the ED, including in the context of an outpatient follow-up visit.

Each expert was asked to rate each identified usability issue’s impact on 2 criteria using a 3-point scale (ie, no impact, some negative impact, or large negative impact). The 2 criteria, selected through discussion and review of the literature, were (1) patient comprehension and (2) patient safety [15,16]. We defined patient comprehension as “the patient’s understanding of the information, for example, what to do next, what to watch for, and what to expect” and patient safety as “the patient’s ability to follow-up and follow-through with recommendations.” As such, patient safety would be negatively impacted by any usability issue that could result in a lack or delay of follow-up, taking the wrong actions, or potential patient harm.

In addition to providing ratings on each criterion for each usability issue, we asked experts to take note of any usability issues that were unclear to them and identify any additional usability issues they may have noticed in the AVS documents that were not identified in stage 1. Each expert’s ratings and notes were then sent back to the research team. One researcher (HJB) reviewed each expert’s ratings and notes for missing data, newly identified usability issues, and any notes of interest. A 30-minute final debrief meeting was scheduled with each expert, wherein researchers (HJB and PC) met with each expert to collect any missing data, ask clarifying questions, and capture any other feedback on the process. Five experts’ ratings and interviews were conducted in August 2020. The final expert’s rating and interview, the geriatrician’s, were conducted in October 2020.

Impact ratings were then converted to a numerical score (0=no impact; 1=some negative impact; 2=large negative impact) for comparison and analysis. Average scores on each criterion were calculated for every usability issue.

**Stage 3: IT Expert Assesses the Likelihood of Addressing Usability Issues**

In the third stage, an electronic health record (EHR) architect from our partner health care organization with extensive institutional knowledge rated each usability issue on the “likelihood we would be able to address it” using a 3-point scale (ie, impossible to address, maybe, or can be addressed). In addition, the IT expert was asked to take note of any comments related to their responses. The expert’s ratings and comments were reviewed by a researcher (HJB) prior to a 30-minute final debrief meeting with researchers (HJB and PC) to discuss ratings and associated comments with the IT expert. Stage 3 was completed in September 2020.

**Ethical Considerations**

This study procedure was exempt from IRB approval as part of a quality improvement initiative. There was no compensation for participation.

**Results**

**Usability Issues and Their Associated Heuristic Violations**

In stage 1, we identified 60 unique usability issues, violating a total of 108 heuristics (each usability issue could violate more than 1 heuristic). We identified violations for each of the categories of heuristics except for 2 heuristics: readability—color and contrast and content—context. The number of violations per heuristic ranged from 0 to 16 (Table 3), with the most frequently violated being clarity of content (16 of 108), absence or lack of information (15 of 108), relevance (13 of 108), and grouping (11 of 108).

In stage 2, clinical, patient, and care partner experts identified 18 additional usability issues, violating an additional 27 heuristics, including the 2 categories of heuristics not identified in stage 1. The number of violations per heuristic ranged from 0 to 7, with 5 heuristics with no new violations identified by our experts (Table 3).

In total, we identified 78 unique usability issues, violating a grand total of 135 heuristics. The heuristics most frequently violated were absence or lack of information (n=22), clarity of content (n=19), relevance (n=14), and terminology (n=12). All heuristics were violated at least once.
Table 3. Number of heuristic violations identified by stages.

<table>
<thead>
<tr>
<th>Heuristic categories and names</th>
<th>Heuristic violations identified in stage 1 (n=108), n</th>
<th>Heuristic violations identified in stage 2 (n=27), n</th>
<th>Total heuristic violations identified (N=135), n</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Readability</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Color and contrast</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Layout and position</td>
<td>4</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Font and capitalization</td>
<td>5</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>Structure and format</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td><strong>Minimalism</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simple and direct</td>
<td>4</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Progressive level of detail</td>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td><strong>Comprehensibility</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Terminology</td>
<td>10</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>Clarity of headings</td>
<td>5</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td><strong>Content</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clarity of content</td>
<td>16</td>
<td>3</td>
<td>19</td>
</tr>
<tr>
<td>Emphasis</td>
<td>5</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Context</td>
<td>0</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Relevance</td>
<td>13</td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td>Absence or lack of information</td>
<td>15</td>
<td>7</td>
<td>22</td>
</tr>
<tr>
<td><strong>Organization</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grouping</td>
<td>11</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>Order</td>
<td>9</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>Use of subheadings</td>
<td>5</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Navigational tools</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Impact Ratings of Usability Issues

In stage 2, we sought to determine the impact of each usability issue on two criteria: (1) patient comprehension and (2) patient safety. We found that average scores on both criteria ranged from 0 (eg, all experts rated “no impact”) to 1.83 (eg, 5 out of 6 experts rated “large negative impact”). The highest rated usability issues included, for example, that “there [was] no indication as to whether the medication list [was] up-to-date, or even if it was reviewed by the ED” (Table 4). This issue scored 1.5 on the patient comprehension criterion and 1.67 on the patient safety criterion. Additional examples are included in Table 4.

Further, we wanted to see if there were differences between the experts’ impact ratings. It was found that on average the older adult care partner used the rating “large negative impact” more frequently than the clinical experts—for example, 37 times when rating usability issues on patient comprehension; the next most used being 23 times (Table 5). Finally, a significant correlation between our 2 criteria, patient comprehension and patient safety, were identified but not between any participant ratings (eg, there was no significant correlation between the 2 ED physicians on either criterion).
Table 4. Highest rated usability issues, the heuristics they violate, their average impact scores on patient comprehension and patient safety, and their likelihood of being addressed.

<table>
<thead>
<tr>
<th>Highest-rated usability issues</th>
<th>Heuristics violated</th>
<th>Average impact score on patient comprehension</th>
<th>Average impact score on patient safety</th>
<th>Likelihood of being addressed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The section “what’s next” is similar to the “instructions” section and presents conflicting information from what is listed under “instructions.” It is unclear to what extent the “what’s next” section relates to the “follow-up” section.</td>
<td>Terminology • Simple and direct • Grouping • Clarity of headings</td>
<td>1.83</td>
<td>1.83</td>
</tr>
<tr>
<td>2</td>
<td>The first page of the AVS&lt;sup&gt;a&lt;/sup&gt; document is cluttered and the information is not presented in a way that makes sense.</td>
<td>Use of subheadings • Progressive level of detail • Grouping</td>
<td>1.83</td>
<td>1.5</td>
</tr>
<tr>
<td>3</td>
<td>AVS is written at a high comprehension level. No visuals or graphics to support comprehensibility. No contact for services that could support people with low reading comprehension (eg, cognitive impairments and nonnative English speakers)</td>
<td>Absence or lack of information • Simple and direct</td>
<td>1.67</td>
<td>1.67</td>
</tr>
<tr>
<td>4</td>
<td>No instructions to follow-up to have wound checked or stitches removed (or who to do this with). The only follow-up mentioned is with rehab and they are not going to do this.</td>
<td>Absence or lack of information • Context</td>
<td>1.67</td>
<td>1.67</td>
</tr>
<tr>
<td>5</td>
<td>The “what’s next” section needs to include a list of the tasks that the patient needs to do next. It should also be grouped with “follow up.”</td>
<td>Absence or lack of information • Grouping</td>
<td>1.5</td>
<td>1.67</td>
</tr>
<tr>
<td>6</td>
<td>There is no indication as to whether this medication list is up-to-date, or even if it was reviewed by the ED.&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Context • Absence or lack of information</td>
<td>1.5</td>
<td>1.67</td>
</tr>
</tbody>
</table>

<sup>a</sup>AVS: after visit summary.
<sup>b</sup>ED: emergency department.

Table 5. Average impact scores and the number of highly rated usability issues by experts.

<table>
<thead>
<tr>
<th>Patient comprehension (n=76 usability issues)</th>
<th>Oldest adult care part- ner representative</th>
<th>Nurse with transition-al care expertise</th>
<th>Geriatrician&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Emergency medicine physician 1</th>
<th>Emergency medicine physician 2</th>
<th>ED&lt;sup&gt;b&lt;/sup&gt; nurse</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average impact score</td>
<td>1.197</td>
<td>1.080</td>
<td>1.026</td>
<td>1.184</td>
<td>0.882</td>
<td>0.789</td>
<td>1.026</td>
</tr>
<tr>
<td>Usability issues rated “large negative im- pact” (eg, score=2), n</td>
<td>37</td>
<td>17</td>
<td>23</td>
<td>23</td>
<td>6</td>
<td>9</td>
<td>19</td>
</tr>
</tbody>
</table>

| Patient safety (n=76 usability issues)      | Average impact score                       | 1.276                                 | 1.120                   | 0.961                          | 0.816                | 0.421                | 0.645   | 0.872 |
|---------------------------------------------|--------------------------------------------|---------------------------------------|-------------------------|--------------------------------|----------------------|----------------------|---------|
| Usability issues rated “large negative im- pact” (eg, score=2), n  | 42                                         | 32                                    | 26                      | 14                             | 3                    | 9                    | 21      |

<sup>a</sup>The geriatrician rated 78 usability issues. All other experts rated 76 usability issues.
<sup>b</sup>ED: emergency department.
Likelihood of Addressing Usability Issues

In stage 3, an IT expert from our partner health system with extensive experience with the ED AVS provided ratings on the “likelihood we would be able to address” each usability issue. Of the 76 usability issues that the expert reviewed, 31 usability issues were rated as “impossible to address,” 21 as “maybe,” and 24 as “can be addressed.” The reasons most cited for being unable to address a usability issue were because the information in the AVS came from an outside vendor (eg, generic patient instructions for wound care) or because the EHR vendor controlled the headers, content, and order of the sections. The reasons cited for why a usability issue may be able to be addressed were because a solution would require additional work for clinicians (eg, ED physicians and nurses) or because it would require an overhaul of the databases that populate the AVS (eg, the name of the clinic to follow-up with). Finally, the usability issues that were most often cited as being able to be addressed were the ones found in sections that the health organization had added to the AVS (eg, generic reminders to wear a seatbelt).

Discussion

Overview

This study found that it is important to integrate diverse expertise to evaluate usability when patient safety is at stake. Twenty-three percent of the identified usability issues (18/78)—a large proportion of which were related to the absence or lack of information—were noted by clinical, patient, and care partner experts in stage 2 and would not have otherwise been identified by HFE experts. The additional 18 usability issues identified by non-HFE experts represent the need to integrate a broader range of expertise.

To conduct a comprehensive heuristic evaluation, expertise from all contexts of use must be considered. In the case of the ED AVS, the experts included (1) the emergency medicine physician who initiates the creation of the AVS in the EHR, but rarely ever sees it printed out; (2) the ED nurse who prints out the AVS and reviews and discusses it with the patient and their care partner upon discharge from the ED; (3) the patient and care partner who receive the document from the ED nurse, carry it home, and who may need to communicate about it with other care partners, family, and their doctor; and (4) the geriatrician (or other primary care doctor) who hears about the ED visit from the patient during their follow-up and may or may not interact directly with the AVS. Thus, the usability of the AVS may differ between the multiple distinct contexts of use. Methods that capture the complex and, on occasion, conflicting perspectives of relevant experts are required to appropriately assess usability and inform redesign.

Similar to findings from a study comparing clinician and patient ratings of nonroutine events, our results demonstrate discrepancies in the impact ratings of different experts [17]. Particularly, the older adult care partner rated usability issues as having a more negative impact on patient comprehension and patient safety. The scores from the geriatrician and nurse with transitional care expertise were similarly high, which may point to poorer usability of the AVS in post ED discharge contexts [18]. Including these context-specific experts in evaluating the impact of the identified usability issues aligns the design priorities with the experience of patients and their care partners upon leaving the ED. Aligning design priorities with the experience of patients and their care partners is a key aspect of designing patient-centered systems [19].

These initial steps at capturing a variety of context-specific expertise point to a unique challenge: How do we integrate these perspectives and choose where to focus our design efforts? This reconciliation of multiple perspectives is a pervasive challenge for diverse health care design teams [20]. One way to address this is by clearly defining an aim, for example, design a patient-centered discharge process, that can guide the integration and prioritization of perspectives in a design team with representation from multiple stakeholders. HFE methods such as participatory design and co-design offer frameworks for doing this [21-24].

The 3-staged method introduced in this paper also begins to bridge the gap between heuristic evaluation and redesign. Capturing insight from an IT expert in stage 3 about what it would take to address each identified usability issue provides practical feedback that can be incorporated into a redesign process. Further, an EHR architect, in particular, may provide insight into the level at which each usability issue could be addressed, for example, at the health system level or at the level of the EHR vendor. By engaging IT during the evaluation of the AVS versus later in the design process, resources can be used more efficiently. Furthermore, given the challenges, frontline staff must upskill well-designed, usable technologies; this method may also bridge the gap from redesign to implementation by avoiding designing a solution that cannot be implemented [25].

Lessons Learned

Our staged method for heuristic evaluation produced uniquely practical insight while remaining efficient. The staged approach allowed for the combined benefit of group heuristic evaluation, that is, the inclusion of multiple HFE experts during initial usability issue identification and the efficient solicitation of feedback from stakeholders with their unique expertise.

Time Investments

The 7 non-HFE experts whose feedback was obtained in stages 2 and 3 spent between 1.5 and 2 hours in total reviewing the usability issues on their own and then debriefing with 2 HFE experts. Four of the HFE experts contributed solely to the stage 1 meeting, that is, 1.5 hours of their time; 2 additional HFE experts contributed an additional 2-3 hours of support in taking notes and preparing an initial report of usability issues after the stage 1 meeting. The remaining 2 HFE experts were heavily involved in the preparation for and execution of all 3 stages, for example, communicating and scheduling with experts, reviewing expert’s feedback, debriefing, and so forth.

Role of HFE Experts

Given our staged approach, HFE experts played different roles at different points in time. During stage 1, HFE experts were the main source of identifying usability issues and assessing
which heuristics those issues violated. During stages 2 and 3, HFE experts served more as facilitators to capture insights from other non-HFE experts and translate them into usability issues, heuristic violations, and relevant feedback on our ability to address those issues.

Selection of Experts
An important aspect of this study is the selection of experts who have relevant context-specific expertise. For example, to represent the interest of a primary care doctor who would follow-up with a patient post ED visit, we selected a geriatrician who is likely to see patients from the population we are designing for, that is, older adults (65+ years) with a recent fall or urinary tract infection. Further, in selecting the IT expert for stage 3, their extensive experience with the ED AVS, as in, how it has been changed over time by the EHR vendor and by the health system, and the processes through which it gets changed within the health system, was essential to providing useful data.

Limitations
A few limitations of this study should be noted. First, given this study was not designed to be generalizable, we used small sample sizes, for example, 6 experts that provided feedback during stage 2. Future work could more extensively explore the discrepancies between experts’ perspectives by increasing the sample size. These data may alter how relevant one considers a single type of expert’s perspective to be, for example, if there is little significant difference between certain experts. Particularly, additional patient and care partner perspectives may be warranted to capture the variety of experiences patients have based on their identity, cognitive abilities, living situation, and so forth.

Conclusions
Capturing relevant context-specific expertise in heuristic evaluation results in more comprehensive identification of usability issues and their impacts. Despite being challenging to integrate, experts’ unique perspectives must be considered to design patient-centered systems. A staged approach to heuristic evaluation may be a useful tool to more reliably identify usability issues that are significant in the patient experience and translate those into actionable redesign.

Acknowledgments
This research was supported by grant 5R18HS026624 from the Agency for Healthcare Research and Quality, and Clinical and Translational Science Award program (grant UL1TR002373), through the NIH National Center for Advancing Translational Sciences. The project was also supported by the University of Wisconsin School of Medicine and Public Health through Wisconsin Partnership Program funding to the Wisconsin Institute for Healthcare Systems Engineering and the Health Innovation Program. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality and the NIH.

Data Availability
Data from this study are not publicly available.

Conflicts of Interest
None declared.

References


Abbreviations

AVS: after visit summary
ED: emergency department
EHR: electronic health record
HFE: human factors engineering

https://humanfactors.jmir.org/2023/1/e43729 JMIR Hum Factors 2023 | vol. 10 | e43729 | p.312 (page number not for citation purposes)
Evaluating the Usefulness and Ease of Use of a Next-Generation–Connected Drug Delivery Device for Growth Hormone Therapy: Qualitative Study of Health Care Professionals’ Perceptions

José I Labarta¹,², MD; Paul Dimitri³, BSc, MBChB, PhD; Matthew Keiser⁴, BS, MS; Ekaterina Koledova⁵, MD, PhD; Octavio Rivera-Romero⁶, PhD

¹Unit of Endocrinology, Department of Pediatrics, Hospital Universitario Miguel Servet, Zaragoza, Spain
²Instituto de Investigación Sanitaria Aragón, Zaragoza, Spain
³Department of Paediatric Endocrinology, Sheffield Children's NHS Foundation Trust, Sheffield, United Kingdom
⁴Ares Trading SA (an affiliate of Merck KGaA), Eysins, Switzerland
⁵Global Medical Affairs Cardiometabolic & Endocrinology, Merck Healthcare KGaA, Darmstadt, Germany
⁶Electronic Technology Department, Universidad de Sevilla, Seville, Spain

Corresponding Author:
Octavio Rivera-Romero, PhD
Electronic Technology Department
Universidad de Sevilla
Avda. Reina Mercedes s/n
Seville, 41012
Spain
Phone: 34 954 55 27 89
Email: orivera@us.es

Abstract

Background: Digital solutions targeting children’s health have become an increasingly important element in the provision of integrated health care. For the treatment of growth hormone deficiency (GHD), a unique connected device is available to facilitate the delivery of recombinant human growth hormone (r-hGH) by automating the daily injection process and collecting injection data such that accurate adherence information is available to health care professionals (HCPs), caregivers, and patients. The adoption of such digital solutions requires a good understanding of the perspectives of HCPs as key stakeholders because they leverage data collection and prescribe these solutions to their patients.

Objective: This study aimed to evaluate the third generation of the easypod device (EP3) for the delivery of r-hGH treatment from the HCP perspective, with a focus on perceived usefulness and ease of use.

Methods: A qualitative study was conducted, based on a participatory workshop conducted in Zaragoza, Spain, with 10 HCPs experienced in the management of pediatric GHD from 7 reference hospitals in Spain. Several activities were designed to promote discussion among participants about predefined topics based on the Technology Acceptance Model and the Unified Theory of Acceptance and Use of Technology to provide their perceptions about the new device.

Results: Participants reported 2 key advantages of EP3 over previous easypod generations: the touch screen interface and the real-time data transmission functionality. All participants (10/10, 100%) agreed that the new device should be part of a digital health ecosystem that provides complementary functionalities including data analysis.

Conclusions: This study explored the perceived value of the EP3 autoinjector device for the treatment of GHD by HCPs. HCPs rated the new capabilities of the device as having substantial improvements and concluded that it was highly recommendable for clinical practice. EP3 will enhance decision-making and allow for more personalized care of patients receiving r-hGH.

(JMIR Hum Factors 2023;10:e46893) doi:10.2196/46893
Introduction

Background

Digital solutions targeting children’s health outcomes are rapidly gaining traction in health care [1]. The World Health Organization introduced digital health as a broad umbrella term encompassing eHealth, which refers to the design and use of information and communication technologies to support the promotion, prevention, treatment, and maintenance of health outcomes [2]. With the introduction and implementation of new digital health solutions, there needs to be an understanding of how these technologies can best be implemented within clinical care pathways; in the patient’s home; and the broader environment in a way that supports connectivity locally, regionally, and globally. eHealth includes mobile health (mHealth)—health-related services delivered via mobile communications devices [3]—which allows health care services to be accessed and delivered remotely in real-world settings. This enables more accurate real-time collection of a large amount of data about health conditions and behaviors [4-6], using advanced analytical techniques to assess, for example, adherence and the effects of treatment on clinical outcomes; these data can be collated at an individual or population-based level. Communication, education, social participation, and treatment reminders are other examples of how mobile-enabled health care services can be used. Such technological developments are triggering a paradigm shift from standard face-to-face interventions toward a more integrated, patient-centered, personalized, and potentially more cost-effective health care approach. mHealth has the potential to improve treatment outcomes and patients’ quality of life [7], as shown by the use of SMS text message reminders to improve medication adherence and perceived quality of life in adolescents with asthma [8] and digitally enabled continuous glucose monitoring in children and adolescents with type 1 diabetes mellitus [9].

Developing Digital Health Tools

Designing digital health tools for children and adolescents requires specific considerations that relate to the anatomical, physiological, and psychosocial changes during their development [10]. These include changes in children’s developmental characteristics as they mature, parent-child dynamics, and the transition of children into adult health care [11]. Type 1 diabetes mellitus is a good example of where it is particularly important to effectively manage the transition from pediatric to adult health care [12], during which technologies can play a fundamental role [9,13,14]. An approach to developing digital tools to support pediatric health care is to integrate user-centered design (UCD). UCD is an evidence-based framework informed by the needs and understanding of specific user groups at every stage of the design process and is invaluable in the development of mHealth technologies [15,16]. It is part of the International Organization for Standardization (ISO) standard “Ergonomics of human-system interaction—Part 210: Human-centered design for interactive systems” [17] and is endorsed by the World Health Organization [18]. UCD aims to create solutions that meet the specific needs, characteristics, preferences, and tasks of the intended users [17,19]. Systems developed using an iterative design process following UCD principles are easy to use and learn, reach high user acceptance and satisfaction levels, and reduce the number of user errors [17,19,20]. Most UCD methods in health care involve service users and service providers in the different stages of the development process [19-21], and involving health care professionals (HCPs) in the development of such solutions may have a positive impact on their perceived reliability [22,23]. Despite the apparent value of UCD, a systematic review of 69 randomized studies of mobile apps designed to support patients with chronic diseases reported that robust developmental factors are rarely adopted during the design stage, with approximately only one-third of the studies reporting user or HCP involvement [24]. Examples of where UCD was applied to the development of digital health solutions for pediatric health care that did involve HCPs include the mHealth tool, the Pictorial Support in Person-Centered Care for Children app, and the development of an electronic cross-facility health record for pediatric palliative care [25,26].

Digital Health Tools for Growth Hormone Deficiency

Digital health tools have been used to support patients in the self-management of pediatric endocrine disorders, such as growth hormone deficiency (GHD). Long-term management of GHD is often challenging for children, their caregivers, and HCPs, as treatment requires daily injections over many years, either self-administered or administered by caregivers [27]. Connected medical devices can be used to facilitate this process by automating the injection, delivering accurate predetermined doses, improving comfort, and reducing injection-related anxiety. Using these devices to monitor therapy by digitally recording daily injections can improve adherence to such long-term therapy through the early detection of suboptimal adherence and, therefore, appropriate intervention by HCPs. Poor adherence can lead to reduced efficacy, increased comorbidities, and increased health care costs and has long been associated with growth hormone (GH) treatment and thus underpins the need and value of objectively measuring adherence by a connected device to drive early intervention [28]. Currently, there is only 1 digitally enhanced, connected autoinjector available to deliver recombinant human GH (r-hGH; somatropin [Saizen], the health care business of Merck KGaA, Darmstadt, Germany) treatment—the easypod (the health care business of Merck KGaA, Darmstadt, Germany) device, which has, so far, been approved in >40 countries. This device has been widely used in pediatric research and practice to improve treatment adherence [29] by facilitating the collection of real-time injection data, so that reliable, accurate information about adherence to treatment is available to HCPs for assessment. Furthermore, population data from these devices...
provide a means of developing prediction tools to support clinical decision-making [30]. As users and prescribers of new digital health technologies to support pediatric growth therapies, it is important to garner HCPs’ perspectives about the acceptance of these devices during their design and development to test usability, feasibility, and acceptability; this was the rationale for conducting this study. Several qualitative studies exploring HCPs’ perceptions about factors and barriers related to digital health acceptance in endocrinology and other therapy areas have been published in the scientific literature [29,31-34]. Some have explored HCPs’ perceptions about mHealth tools used in children’s health care [29,31-37], concluding that early engagement of end users is critical to the development and effective implementation of such tools to enhance patient-centric care. Notably, a mixed methods formative research study has explored technology acceptance and the use of digital health tools for the emotional support of parents of children undergoing GH treatment, using educational content to help parents and caregivers understand their children’s treatment journey [38]. However, to the best of our knowledge, no study has explored HCPs’ perceptions about the acceptance of mHealth solutions (and their technological evolution) to support pediatric GH therapy. For example, digital interventions based on recorded adherence data have been implemented in the context of r-hGH treatment [39].

The third generation of the easypod device (EP3) was designed with patients and caregivers in mind; however, evaluations using UCD methods to better understand the HCPs’ perspective to support the implementation and acceptance of the device in relation to their specific needs (eg, by better understanding the barriers to implementation and advantages of the device) were not performed during the development phase.

**Objectives**

Therefore, to add to the existing published literature and to build upon the insights from previous studies, this study was conducted to assess 2 constructs of technology acceptance of EP3—perceived usefulness and ease of use—compared with the current digital health device used to support and deliver pediatric r-hGH therapy from an HCP perspective.

**Methods**

**Methodological Models**

Several Technology Acceptance Models (TAMs) and theories have been developed to explain the intention to use technological solutions [40-47]. As an example, the TAM is a behavioral model of user acceptance of technology that has been widely used in research [40]. It posits that the perceived usefulness and perceived ease of use of a digital solution predict the intention to use it and, therefore, its actual use. Several versions of TAM have been developed incorporating additional factors such as social norms [43,44]. As this study aims to explore how the technological advances could have an impact on HCP perspectives, only the core factors that are directly related to the technology being assessed have been considered (ie, perceived usefulness and ease of use). In this regard, perceived usefulness is defined as “the degree to which a person believes that using a particular system would enhance his or her job performance.” whereas perceived ease of use is defined as “the degree to which a person believes that using a particular system would be free of effort.”

The Unified Theory of Acceptance and Use of Technology (UTAUT) [45] identifies 4 main constructs that play a significant role as direct determinants of user acceptance and use behavior: performance expectancy, effort expectancy, social influence, and facilitating conditions. The first 2 are related to the abovementioned TAM’s constructs. Performance expectancy is defined as “the degree to which an individual believes that using the system will help him or her to attain gains in performance.” According to UTAUT, this construct is the strongest predictor of intent to use. It is directly related to perceived usefulness defined in the TAM. Effort expectancy is defined as “the degree of ease associated with the use of the system” and encapsulates the same concept as the TAM’s perceived ease of use construct.

In the health domain, Kim and Park [42] developed the Health Information TAM (HITAM). This model expands upon the TAM by adding specific factors related to health. Perceived usefulness and perceived ease of use are still considered as significant mediators of user’s attitude, which directly influences the behavioral intention and, hence, use. An additional core construct is also included in HITAM, namely, perceived threat, which is derived from the Health Belief Model (HBM) [48]. The HBM is a social cognition model used to explain health behavior change. It suggests that belief in a personal threat, together with belief in the effectiveness of the proposed behavior, predicts the likelihood of engaging in that behavior.

Finally, Wang et al [47] defined a model that integrates UTAUT and Task-Technology Fit (TTF) [49] to understand how consumers accept health care wearable devices. TTF posits that “for a technology to have a positive impact on individual performance, the technology: (1) must be utilised and (2) must be a good fit with the tasks it supports” [49]. This model incorporates components derived from TTF (technology characteristics, task characteristics, and TTF) to UTAUT. Wang et al [47] found that performance expectancy was the most important determinant of behavioral intention. They also determined that technology characteristics could positively predict effort expectancy, whereas TTF directly influenced behavioral intention through the mediating role of performance expectancy.

In this study, we focused on 2 constructs included in the TAM—perceived usefulness and perceived ease of use—which are also components of the UTAUT, HITAM, and HBM. TAM and UTAUT are general acceptance models and could be used as the basis for studies in any domain; however, HITAM and HBM are models specifically defined for the health domain. Therefore, as our study focused on these 2 main constructs, the theoretical foundations from both general acceptance and health-related models were valid to explore how pediatric HCPs perceive the potential impact of technology evolution on the acceptance of an mHealth device, namely, easypod.
**Study Design**

This theory-driven qualitative study was conducted through a participatory workshop involving 10 HCPs (n=6, 60% endocrinologists; n=2, 20% nurses; and n=2, 20% pharmacists), with the workshop lasting 3 hours (Figure 1). Several predefined questions designed based on perceived usefulness and perceived ease of use were discussed during the workshop session.

**Figure 1.** Photos of participants taken during the workshop sessions.

---

**Study Cases**

Easypod is the only digitally enhanced, connected autoinjector device available to deliver r-hGH treatment. Therefore, we used 2 versions of this device as the comparable study cases: the current easypod device (EP2) and EP3 that is currently in development to deliver r-hGH and monitor real-time adherence to therapy (Figure 2).

Both devices record the date, time, and dose received, but EP2 cannot transmit these data until the user or carer places it on a separate docking station and activates transmission. In contrast, EP3 transmits the data automatically, with no requirement for user activation or a separate docking station for data transmission.

EP3 is taller and slimmer than EP2 and has a removable and rechargeable battery; a large, easier-to-read touch screen; and a skin sensor with 360° coverage, enabling improved skin contact compared with the 180° coverage with EP2; thus, it is intended to make injections easier and more accurate. The injection button on EP3 is at the front of the device, whereas on EP2, it is at the top. The needle is hidden on both devices to minimize needle phobia and patient anxiety, with the additional feature of automated needle detachment on EP3. The comfort settings (injection speed; injection depth; and needle speed, which can be adjusted by an HCP according to patient preference; and injection time, ie, the duration for which the needle remains in the skin) are a feature shared by the 2 devices (Figure 2).

Regarding safety, EP3 will comply with all the latest and relevant standards for medical devices (ISO 11608, International Electrotechnical Commission 60601, and ISO 62304).
Figure 2. Attributes of the current easypod device (EP2) and the third generation of the easypod device (EP3).

Study Setting
The participatory workshop, upon which this qualitative study was based, was conducted at the University Hospital Miguel Servet, Zaragoza, Spain, on February 23, 2022, in 2 meeting rooms situated in proximity on the same floor of the hospital. Both rooms had the necessary technical equipment (video projector, audio system, and computer) required for the sessions. The workshop was conducted in Spain, as there is a representative sample of physicians who have used digital health solutions available in clinical practice there.

Study Participants
Participants included HCPs from 7 different hospitals from 6 different regions in Spain, with experience in the management of growth disorders using r-hGH treatment in pediatric patients, either with or without previous experience of using EP2. The management of GH therapy in Spain involves multidisciplinary teams comprising pediatric endocrinologists, nurses, and pharmacists. As such, we sought views from representatives of these disciplines about the usefulness and ease of use of EP2 and EP3, ensuring that the sex and expertise of participants (endocrinologist, nurse, or pharmacist) were considered when selecting the final sample of participants. Participants were grouped into 2 teams, balanced in terms of their professional expertise and sex, and each team was led by a facilitator. Both facilitators had experience of conducting qualitative studies.

Ethics Approval and Informed Consent
Ethics approval for this study was obtained from the ethics committee of research of the Universidad de Sevilla (ID 2593-N-21). Participants’ agreement for participation was obtained through an informed consent process.

Participatory Workshop

Workshop Design
The participatory workshop consisted of several activities in which participants discussed a set of predefined topics. The workshop was designed by 2 experts in participatory health informatics. Several procedures and materials were designed to create an appropriate working environment to facilitate discussions. The topics to be discussed incorporated the 2 constructs of the TAM (perceived usefulness and perceived ease of use) and were led by a multidisciplinary team including experts in digital health, participatory health informatics, and technology acceptance and HCPs. Before beginning each activity, the facilitator provided clear instructions to participants about how to perform the activity, and any questions were resolved. These activities were grouped into 5 phases (Figure 3).
Phase 1

Phase 1 aimed to briefly introduce the project; its objectives, facilitators, and phases; and specific tasks to be performed during the workshop. All participants were in the same room and the facilitator asked them to briefly introduce themselves. We also clarified technical terms such as “digital health solution” or “device,” and any questions were resolved. Finally, participants were grouped into 2 teams, and each team performed the activities in a separate room.

Phase 2

Phase 2 consisted of 2 activities (activities 1 and 2) that were performed independently by each team. During activity 1, participants discussed the ergonomics of the packaging (the cases in which the devices are held and transported) of each device (EP2 and EP3). The facilitators provided participants with the packaging for each device. Participants spent a short time inspecting the cases, trying to open and close them, and compared their physical characteristics. Any questions were resolved. The facilitator then asked participants to highlight the strengths and weaknesses of the packaging for the 2 devices, considering 2 scenarios: from the user and the caregiver perspectives.

Activity 2 focused on the ergonomics of the 2 devices. At the beginning of this activity, the facilitators provided participants with a prototype of each device, the dimensions and weights of which were equal to the commercial devices but did not implement all their software functionalities. Participants spent a short time inspecting the devices, performed a simulated injection, and compared their physical characteristics. In addition, an introductory video presenting the main characteristics of each device was played, and any questions were resolved. The facilitators then provided the teams with an activity 3 template to prompt discussions and a set of sticky cards that represented the topics predefined for this activity (an example of which is shown in Figure 4). Each sticky card comprised short text and imagery representing a predefined topic. Next, participants randomly selected a topic card and stuck it on to the template. Participants discussed the selected topic and highlighted the strengths and weaknesses of both devices. A participant summarized the opinions reported by the team, making brief notes on the template. Once the discussions were completed, a new topic was randomly selected, and the same procedure was repeated until all predefined topics were discussed or the time to complete the activity was reached. Some examples of the predefined topics for this activity were the appropriateness of device weight, dimensions, screen location, administration button size, administration button location, and feedback light location.

Figure 3. Phases of the workshop.

Figure 4. An example of one of the sticky cards representing the topics predefined for activity 2.
**Phase 3**

The objective of phase 3 was to explore participants’ opinions about the differences and similarities of both devices when they are used to manage treatments and the potential impact on the technology acceptance perceived by participants. The complete process to perform treatment administration using the easypod devices was split into 6 main tasks: dose configuration, cartridge replacement, needle detachment and attachment, performing injections, data transmission, and providing user feedback. We designed an activity for each task (activities 3-7) that was independently performed by each team. An activity 3 template and a set of sticky cards representing the predefined topics were designed for each activity. The procedures followed for each activity were the same as those used in activity 2. First, a short video explaining how to perform the corresponding task using both devices was played, and any questions were resolved. Next, a topic was randomly selected and discussed among participants. Participants’ opinions were noted down on the template during discussions. The process of selecting a topic and discussing it was repeated until the time to complete the activity was reached or until all predefined topics were discussed. Some examples of the predefined topics for these activities were perceived ease of performing the task, perceived safety while performing the task, potential human errors, perceived ease of interaction with the device, perceived ease of teaching the task to HCPs, patients, and caregivers, perceived effort required to perform the task, and potential benefits of using the device.

**Phase 4**

Phase 4 began with an individual activity (activity 8). In this activity, each participant individually completed a questionnaire. The main idea of this individual activity was to give participants the opportunity to report any opinion that they did not provide during the previous activities. The questionnaire consisted of 25 items aimed to assess the participants’ perceptions about how the improvements included in EP3 influenced acceptance in terms of 4 domains: perceived usefulness, perceived ease of use, ease of learning, and intention to use and recommend. Each item of the questionnaire consisted of 2 components: a 5-point Likert question and an open-ended question. The first component aimed to present the specific question to be discussed and to allow participants to think about its perceived relevance. The second component asked participants to justify the score assigned. Although a quantitative questionnaire was used, we analyzed the data collected in this activity using a qualitative approach. The list of items included in the questionnaire is presented in Multimedia Appendix 1, the results of which were used in a descriptive manner, with the aim of providing new insights and observations that were not otherwise reported during the workshop; mean Likert scores according to domain are presented in Multimedia Appendix 1.

Next, the 2 groups of participants were brought together to present their findings about the most relevant issues reported during the previous activities, potential impact of the improvements included in EP3 on the management of r-hGH treatment, and role of these advances in the broader digital ecosystem. During this phase, we sought to present the perceptions of both teams and encourage the participants from each team to discuss their opinions in depth. Facilitators summarized the most relevant comments reported by participants during the previous activities and then presented these relevant issues separately to prompt discussion among the participants.

**Phase 5**

Finally, during phase 5, some statements representing the most relevant findings were presented to all participants, and facilitators asked the participants to validate these conclusions and gave them the opportunity to add some additional comments to clarify them.

**Data Analysis**

The workshop session was audio recorded. The audio recordings were reviewed by a researcher, after which, relevant comments were transcribed, and information from the facilitators’ notes and text included in the templates were combined into the study data set. Owing to the small sample size of participants, we did not seek to determine the statistical significance or generalizability of the quantitative data collected using the defined questionnaire but, instead, to describe their opinions. In this regard, OR-R reviewed the scores assigned by participants to questions included in the questionnaire and checked whether any additional opinions were provided, to ensure that they were of a qualitative nature. These additional opinions were included in the data set. Then, the data collected in this study were explored qualitatively using an inductive approach following a simplified theory-guided thematic analysis for qualitative data [50]. OR-R reviewed all collected data, coded them, and defined themes, after which, all authors reviewed the proposed themes and refined them until consensus was reached. For this study, the quantitative data were not analyzed.

**Results**

**Overview**

Teams were created from the participating HCPs, considering their professional backgrounds (3/10, 30% endocrinologists; 1/10, 10% nurses; and 1/10, 10% pharmacists in each team). Overall, 5 themes were identified: simplified touch screen interface, real-time data transmission, administration safety, digital ecosystem, and additional improvements (Table 1). All scores reflected conclusions that aligned with the findings from the qualitative data (Multimedia Appendix 1). As an example of the perceived impact of the technological evolution of EP3, HCPs perceived the improvements as having a positive impact on its usefulness and ease of use. These quantitative data showed a high predisposition of the HCPs to use and recommend the new-generation device, demonstrating that they perceived it as an important advancement.
Table 1. The 5 themes identified during the workshop.

<table>
<thead>
<tr>
<th>Themes</th>
<th>Participants’ insights (verbatim)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simplified touch screen interface</td>
<td>“The touch screen is very clear and visible.”</td>
</tr>
<tr>
<td></td>
<td>“More intuitive because of the touch screen.”</td>
</tr>
<tr>
<td></td>
<td>“Bigger screen, better resolution, and touch screen.”</td>
</tr>
<tr>
<td></td>
<td>“Text is easier to read.”</td>
</tr>
<tr>
<td></td>
<td>“Children are more familiar with the use of touch screen.”</td>
</tr>
<tr>
<td></td>
<td>“Bigger screen size makes easy the configuration.”</td>
</tr>
<tr>
<td></td>
<td>“Provides more information in the screen.”</td>
</tr>
<tr>
<td></td>
<td>“The administration button in the frontal location could result in unintended interactions with the touch screen.”</td>
</tr>
<tr>
<td></td>
<td>“New device could cause errors because of the location of the touch screen close to the button.”</td>
</tr>
<tr>
<td></td>
<td>“User could touch accidentally the screen while he/she is administrating his/her treatment.”</td>
</tr>
<tr>
<td>Real-time data transmission</td>
<td>“It is the main improvement, and it is a big advance.”</td>
</tr>
<tr>
<td></td>
<td>“It is a crucial component.”</td>
</tr>
<tr>
<td></td>
<td>“The adherence data transmission is a key factor.”</td>
</tr>
<tr>
<td></td>
<td>“Avoid patients/caregivers to forget completing the process.”</td>
</tr>
<tr>
<td></td>
<td>“Real-time data.”</td>
</tr>
<tr>
<td></td>
<td>“Patients/caregivers have one less action to do.”</td>
</tr>
<tr>
<td></td>
<td>“Independency of family will/skills.”</td>
</tr>
<tr>
<td></td>
<td>“Better adherence monitoring, especially in non-adherent patients.”</td>
</tr>
<tr>
<td></td>
<td>“Automatic data transmission will improve the control of adherence for patient, caregiver and HCP.”</td>
</tr>
<tr>
<td></td>
<td>“Actions from patients/caregivers are not required.”</td>
</tr>
<tr>
<td></td>
<td>“Automatic transmission does not require any action by users.”</td>
</tr>
<tr>
<td></td>
<td>“Data transmission is independent from users.”</td>
</tr>
<tr>
<td></td>
<td>“Patients/caregivers are aware HCPs are accessing data in real-time, therefore this fact will impact positively on their behaviours.”</td>
</tr>
<tr>
<td>Administration safety</td>
<td>“EP3 allows a safety process but EP2 did it [similar EP3 and EP2].”</td>
</tr>
<tr>
<td></td>
<td>“More intuitive.”</td>
</tr>
<tr>
<td></td>
<td>“EP3 is very easy to use.”</td>
</tr>
<tr>
<td></td>
<td>“Clinical settings must use a more secure access technique.”</td>
</tr>
<tr>
<td>Digital ecosystem</td>
<td>“Real adherence data allows to make better decisions.”</td>
</tr>
<tr>
<td></td>
<td>“Improved data usage but care will be the same.”</td>
</tr>
<tr>
<td>Additional improvements</td>
<td>“Removing the needle de-attachment button is a big advantage.”</td>
</tr>
<tr>
<td></td>
<td>“Especially because of the simpler needle de-attachment process.”</td>
</tr>
<tr>
<td></td>
<td>“Especially, EP3 minimises problems because of the needle de-attachment process.”</td>
</tr>
<tr>
<td></td>
<td>“Easier to use and more sophisticated.”</td>
</tr>
<tr>
<td></td>
<td>“EP3 requires less effort for patients [understanding].”</td>
</tr>
<tr>
<td></td>
<td>“Better navigation.”</td>
</tr>
<tr>
<td></td>
<td>“Faster and easier.”</td>
</tr>
<tr>
<td></td>
<td>“Simpler menus.”</td>
</tr>
</tbody>
</table>

Participants were unaware about the functionality of the third generation of the easypod device, whereby, when an injection is performed, the screen is blocked, so that, if touched, nothing happens.

HCP: health care professional.

EP3: third generation of the easypod device.

EP2: current easypod device.

Simplified Touch Screen Interface

The development of a more intuitive interface that improved the clarity and visibility of information displayed and facilitated digital interaction was perceived as important by the participants:

*The EP3 interface is more intuitive; easier to use [and] bigger visual clues.* [Endocrinologist]

*Navigation is faster using touch screen than using buttons.* [Nurse]
example, the simplifications included in the EP3 prototype (fewer steps to configure it or to access the appropriate option, and better navigation) were perceived as making it easier to use, learn, and train. This increased ease of learning and ease of training were considered as valuable by participants because making the device easy to teach and learn has a positive impact on HCPs’ clinical practice. HCPs, especially nurses, support patients and caregivers in the use of these digital devices. In this regard, HCPs introduce the device, explain its use, and resolve any questions from patients and caregivers; this requires time, which is quite limited in their daily practice.

**Real-Time Data Transmission**

Participants repeatedly commented about the importance of collecting adherence data. They considered that data collection should be as transparent as possible for users, reducing the number of additional user interactions, and that the use of a digital solution is crucial for generating a sustainable, trusted, and unbiased adherence data collection method:

- *Measuring adherence is a crucial factor in any disease. The automatic data transmission allows to collect information that currently HCPs do not have access to.* [Endocrinologist]
- *Having real-time data collected automatically allows us to resolve doubts regarding adherence and improve the patient’s management.* [Endocrinologist]

Some HCPs reported their previous experience of using digital solutions in the management of pediatric chronic conditions, in which patients and caregivers did not share their adherence data because additional user actions required to transmit data were perceived as very burdensome. In this regard, the automatic process for real-time data transmission was considered to be a major advantage of the EP3 prototype, as it is transparent and independent of the user and does not require an additional device to transfer data, thus enabling HCPs’ access to real-time data from all patients. This was reported to be extremely valuable by HCPs because the decision-making process can be based on a more realistic data set than the one used previously. Participants, particularly nurses, also commented that training patients and caregivers on how to manage or self-manage pediatric disorders is a key factor, especially when using digital solutions. Participants agreed that the training process would be simplified because no instruction about how to transfer or share data would be required. Patient support programs have been developed to train patients and caregivers in the use of EP2 [17], but some of them do not acquire the appropriate skills or forget the process for transmission, leading to a lack of shared data. Thus, participants agreed that the automatic transmission of data would make the device easier to use, learn, and train. They also stated that features such as tactile interaction and automatic data transmission would both facilitate training and increase the usability of the device. In turn, this could reduce the time and effort required to train patients and caregivers.

Participants also agreed that the availability of real-time adherence data would enable better treatment monitoring and improved decision-making, as automatic data transmission offers a more reliable and realistic data set for both adherent and nonadherent patients, thus avoiding or reducing the current bias caused by the lack of data collected from poorly adherent patients.

**Administration Safety**

Participants considered that administration problems such as false administration or unintended movements during treatment administration could be avoided by the EP3 prototype because of its large contact surface and the 360° skin sensor, which enables better skin contact than the 180° skin sensor in EP2:

- *The EP3 device presents improved processes [needle attachment, cartridge replacement, etc] making it easier to use.* [Nurse]

**Digital Ecosystem**

Despite the advantages of the availability of real-time data, the participants acknowledged that the analysis of such data may increase their workload. As noted by the facilitators, all participants (10/10, 100%) agreed that the new device should be part of a digital health ecosystem that provides complementary functionalities such as data analysis and visualization:

- *I agree [that] new functionalities will be needed. These functionalities must automatically analyse the collected data and send an alarm/warning/alert to the HCP to be addressed.* [Endocrinologist]

Notifications and reminders with educational and motivational content as part of the overall digital health solution were seen as valuable additional elements.

**Additional Improvements**

Relevant participants’ comments about the potential improvements of the case study have been included in this theme:

- *The EP3 [device] could be improved in terms of ergonomics, especially in [terms of] dimensions to be easily transported.* [Nurse]

The main area for improvement reported by participants related to the packaging of EP3; some found it difficult to open and close, adding that the size of packaging could make its transport and storage (in a refrigerator) difficult. Participants also commented that they had expected the EP3 prototype to be much lighter and smaller than EP2.

**Discussion**

**Principal Findings**

In the setting of pediatric GHD, the success of digital solutions—as part of integrated health care—requires an understanding of how HCPs perceive how connected devices facilitate patient management. This qualitative study, involving 10 HCPs from 7 reference hospitals in Spain, provides new information about the perceived usefulness and ease of use of a connected device that has evolved to meet the changing needs of those involved with the management of pediatric GHD. Participants in this study agreed that the new prototype device represents a technological evolution, in that it provides complementary functionalities—including real-time data
Participants reported some negative opinions around ergonomics; size; and storage of EP3, particularly, storage in a refrigerator. In contrast, previous regulatory studies demonstrated that patients and caregivers were satisfied with the size of the device (unpublished data, Emergo by UL; unpublished data, Use-Lab). The increased height of EP3 was a result of a specific design decision to improve the accuracy of the dose administered; this resulted in a tall device but one with improved accuracy. However, ongoing studies to evaluate these factors from a user’s perspective will provide further valuable insights into these issues. Although some participants were concerned about the frontal location of the administration button, this was determined based on the results of human factor studies (unpublished data, Emergo by UL; unpublished data, Use-Lab). The participants’ comments about the need for small dose increments (depending on individual patient requirements based on clinical response and serum insulin-like growth factor–1 levels [54]) to be made available in the device settings were noted, and this is currently being investigated as part of the ongoing development of the EP3 prototype.

Our study presents an evaluation of connected injection devices to deliver r-hGH therapy using a robust methodological approach, the results of which are transferrable to digital health solutions in other therapeutic areas, especially in terms of facilitators for and barriers to technology acceptance. For example, a recent qualitative analysis concluded that barriers and facilitators should be considered for effective implementation of connected health solutions to support children with cancer and their families [35]. Although TAM is sometimes criticized for being very simplistic [55], the aim of our study was not to identify new constructs for TAM but to identify facilitators and barriers related to the core constructs of perceived usefulness and perceived ease of use that are common to other models and theories. These constructs are directly related to the technology being assessed and, therefore, are the most relevant factors for assessing how the technological advances could have an impact on HCP perspectives. However, the authors acknowledge that other frameworks can be used in this regard; for example, the Nonadoption, Abandonment, Scale-up, Spread, and Sustainability framework defines a construct directly related to technology that it is similar to our approach and is related to our findings [56]. Any future studies evaluating the acceptance of EP3 could use the Nonadoption, Abandonment, Scale-up, Spread, and Sustainability framework to explore other constructs such as “Value proposition” and “Adaptors” [56].

**Study Limitations**

A limitation of the study is the fact that it was conducted only in Spain, despite it providing a representative cross-sectional sample of HCPs who have used digital health solutions in clinical practice there. Exploring the perceptions of HCPs regarding EP3 in other countries and regions could be valuable to reflect views in other national and regional health care
systems. Access to r-hGH treatment (from a practical and financial perspective) and advanced digital health solutions (including EP3) is likely to differ between countries. Finally, the small sample size does not allow the generalization of the quantitative data.

Conclusions
This study explored the perceived value of the next-generation EP3 autoinjector device to HCPs, based on their assessment of the device to deliver r-hGH for the treatment of GHD compared with the currently used EP2. HCPs rated the new capabilities of the device, including the large touch screen and automatic data transmission, as substantial improvements. Therefore, this next-generation easypod device, while retaining the key features appreciated by patients such as the hidden needle and comfort settings, has the potential not only to improve and provide a more personalized treatment experience for patients and their caregivers but also to provide real-world and real-time insights for HCPs for improved clinical decision-making.

The overall conclusion of these participants was that the EP3 prototype was highly recommendable, based on their assessment from the viewpoint of HCPs involved in the treatment of growth disorders. It would be valuable to evaluate the perceptions about the usability of EP3 from the patient and caregiver perspective in future studies.

Acknowledgments
The authors would like to thank Luis Fernandez-Luque for his scientific support during the development of this study. Medical writing assistance was provided by Amy Evans of inScience Communications, Springer Healthcare Ltd, the United Kingdom, and was funded by the health care business of Merck KGaA, Darmstadt, Germany. This study was sponsored by the health care business of Merck KGaA, Darmstadt, Germany (CrossRef funder ID 10.13039/100009945).

Data Availability
Any requests for data by qualified scientific and medical researchers for legitimate research purposes will be subject to the data sharing policy of the health care business of Merck KGaA, Darmstadt, Germany. All requests should be submitted in writing to the data sharing portal for the health care business of Merck KGaA, Darmstadt, Germany (https://tinyurl.com/5ary8j58). When the health care business of Merck KGaA, Darmstadt, Germany, has a coresearch, codevelopment, comarketing, or copromotion agreement or when the product has been out-licensed, the responsibility for disclosure might be dependent on the agreement between parties. Under these circumstances, the health care business of Merck KGaA, Darmstadt, Germany, will endeavor to gain agreement to share data in response to requests.

Conflicts of Interest
JIL has participated in advisory boards for the health care business of Merck KGaA, Darmstadt, Germany; Novo Nordisk; and Pfizer. PD has received consultancy fees from the health care business of Merck KGaA, Darmstadt, Germany. MK is a former employee of Ares Trading SA (an affiliate of Merck KGaA), Eysins, Switzerland; current affiliation: EMD Serono, Inc. (an affiliate of Merck KGaA), Rockland, MA, USA. EK is an employee of Merck Healthcare KGaA, Darmstadt, Germany, and holds shares in the company. OR-R has participated in an advisory board for the health care business of Merck KGaA, Darmstadt, Germany, and has received funding from the Universidad de Sevilla and the Ministerio de Universidades of the Spanish Government.

Multimedia Appendix 1
Items included in the Likert survey and mean Likert scores by domain.
[DOCX File, 46 KB - humanfactors_v10i1e46893_app1.docx ]

References


Abbreviations

EP2: current easypod device
EP3: third generation of the easypod device
GH: growth hormone
GHD: growth hormone deficiency
HBM: Health Belief Model
HCP: health care professional
HITAM: Health Information Technology Acceptance Model
ISO: International Organization for Standardization
mHealth: mobile health
r-hGH: recombinant human growth hormone
TAM: Technology Acceptance Model
TTF: Task-Technology Fit
UCD: user-centered design
UTAUT: Unified Theory of Acceptance and Use of Technology
and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Introduction to the Coproduction of Supervision Standards for Digital Peer Support: Qualitative Study

Caroline Collins-Pisano1*, BA; Michael Johnson2*, MA; George Mois3*, LICSW, PhD; Jessica Brooks4*, PhD; Amanda Myers5*, MPH; Deanna Mazina6*, MPH; Marianne Storm7*, PhD; Maggie Wright8*, CPS; Nancy Berger9*, CPS; Ann Kasper10*, CPS, MA; Anthony Fox11*, CRSS; Sandi MacDonald12*, CPS; Sarah Schultze13*, BA; Andrew Bohm14*, MS, PhD; Julia Hill15; Karen Fortuna16*, LICSW, PhD

1Department of Psychiatry, Dartmouth College, Hanover, NH, United States
2The Commission on Accreditation of Rehabilitation Facilities International, Tucson, AZ, United States
3College of Applied Health Sciences, University of Illinois, Champaign, IL, United States
4Department of Psychiatry, School of Medicine and Public Health, University of Wisconsin–Madison, Madison, WI, United States
5The Heller School for Social Policy and Management, Brandeis University, Waltham, MA, United States
6Watertown Health Department, Watertown, MA, United States
7Faculty of Health Sciences and Social Care, Molde University College, Molde, Norway
8Families in Trauma and Recovery, Fife, United Kingdom
9Roxbury Community College, Boston, MA, United States
10Kasper Connects, Portland, OR, United States
11Tennessee Mental Health Consumer's Association, Nashville, TN, United States
12International Association of Pre-Menstrual Disorders, Boston, MA, United States
13School of Social Work, University of New Hampshire, Durham, NH, United States
14The Dartmouth Institute for Health Policy and Clinical Practice, Geisel School of Medicine, Dartmouth College, Hanover, NH, United States
15Dartmouth College, Hanover, NH, United States
16Department of Psychiatry, Geisel School of Medicine, Dartmouth College, Concord, NH, United States
*these authors contributed equally

Corresponding Author:
Karen Fortuna, LICSW, PhD
Department of Psychiatry
Geisel School of Medicine
Dartmouth College
2 Pillsbury Street
Suite 401
Concord, NH, 03301
United States
Phone: 1 603 722 5727
Email: karen.l.fortuna@dartmouth.edu

Abstract

Background: Digital peer support enhances engagement in mental and physical health services despite barriers such as location, transportation, and other accessibility constraints. Digital peer support involves live or automated peer support services delivered through technology media such as peer-to-peer networks, smartphone apps, and asynchronous and synchronous technologies. Supervision standards for digital peer support can determine important administrative, educative, and supportive guidelines for supervisors to maintain the practice of competent digital peer support, develop knowledgeable and skilled digital peer support specialists, clarify the role and responsibility of digital peer support specialists, and support specialists in both an emotional and developmental capacity.

Objective: Although digital peer support has expanded recently, there are no formal digital supervision standards. The aim of this study is to inform the development of supervision standards for digital peer support and introduce guidelines that supervisors can use to support, guide, and develop competencies in digital peer support specialists.
Methods: Peer support specialists that currently offer digital peer support services were recruited via an international email listserve of 1500 peer support specialists. Four 1-hour focus groups, with a total of 59 participants, took place in October 2020. Researchers used Rapid and Rigorous Qualitative Data Analysis methods. Researchers presented data transcripts to focus group participants for feedback and to determine if the researcher’s interpretation of the data match their intended meanings.

Results: We identified 51 codes and 11 themes related to the development of supervision standards for digital peer support. Themes included (1) education on technology competency (43/197, 21.8%), (2) education on privacy, security, and confidentiality in digital devices and platforms (33/197, 16.8%), (3) education on peer support competencies and how they relate to digital peer support (25/197, 12.7%), (4) administrative guidelines (21/197, 10.7%), (5) education on the digital delivery of peer support (18/197, 9.1%), (6) education on technology access (17/197, 8.6%), (7) supervisor support of work-life balance (17/197, 8.6%), (8) emotional support (9/197, 4.6%), (9) administrative documentation (6/197, 3%), (10) education on suicide and crisis intervention (5/197, 2.5%), and (11) feedback (3/197, 1.5%).

Conclusions: Currently, supervision standards from the Substance Abuse and Mental Health Services Administration (SAMHSA) for in-person peer support include administrative, educative, and supportive functions. However, digital peer support has necessitated supervision standard subthemes such as education on technology and privacy, support of work-life balance, and emotional support. Lack of digital supervision standards may lead to a breach in ethics and confidentiality, workforce stress, loss of productivity, loss of boundaries, and ineffectively serving users who participate in digital peer support services. Digital peer support specialists require specific knowledge and skills to communicate with service users and deliver peer support effectively, while supervisors require new knowledge and skills to effectively develop, support, and manage the digital peer support role.

(Keywords: digital peer support; mHealth; standards; guideline; peer support; supervision; focus group; qualitative data analysis; competency; competencies; health care education; professional education; professional development; continuing education)

Introduction

Digital peer support may show longevity past the COVID-19 era. In April 2020, a survey of 180 peer support specialists from 23 states found a 95% increase in peer support specialists offering digital peer support and a 90% increase in peer support specialists’ confidence in digital peer support [1]. Peer support has transformed to be offered through digital technologies and telemental health sessions [1]. Digital peer support involves live or automated peer support services delivered through technology media such as peer-to-peer networks, smartphone apps, and asynchronous and synchronous technologies [2]. Peer support has been described as social-emotional support, frequently coupled with instrumental support [3]. It is often provided by persons with a lived experience of a mental health condition or substance use disorder to others sharing a similar mental health condition and substance use disorder or mutually offered between both people. The World Health Organization defines peer support as an essential mental health service [4].

Digital peer support expands the reach and practices of in-person peer support and enhances service users’ abilities to engage in mental and physical health services despite barriers such as location, transportation, and other accessibility constraints. Digital peer support sessions have no geographic or time limitations, promote high levels of engagement when developed with peer support specialists as partners, engage service users in mental health services outside of clinical settings, and have access to harder-to-reach groups such as rural residents, older adults, and people experiencing homelessness [5]. Like in-person peer support, digital peer support enhances the quality of life, functioning, social support, recovery, hope, and empowerment. Studies on the feasibility and preliminary effectiveness have found that digital peer support services reduce mental health symptoms and promote engagement in services [5].

Although digital peer support has gained traction globally, at present, no formal digital supervision standards have been put in place. Supervision is considered critical for the development of competent mental health workers [6]. Supervision standards have the potential to aid in the transition to telemental health delivery and help telehealth workers to develop the competencies needed for the telemental health services [6]. According to the Substance Abuse and Mental Health Services Administration (SAMHSA), supervision is a collaborative activity between a supervisor and their workers in which the supervisor guides and supports the worker to promote the fidelity-adherent delivery of services and the development of the worker’s skills and knowledge surrounding the peer support [7]. SAMHSA is an agency within the United States Department of Health and Human Services, which leads public health efforts to advance the behavioral health of the United States.

Supervision of peer support specialists can help to enhance problem-solving skills, improve clarity in the decision-making process, empower workers, increase satisfaction, and help peer support specialists to deliver better services due to opportunities for reflection and discussion of work and work-related issues [8]. Effective supervision practices can also help mental health organizations to better manage resources, improve performance, and increase morale [8].

The transition to, and maintenance of, digital peer support offers challenges for both peer support specialists and service users. For example, a recent study about combining web-based and offline peer support discusses challenges related to the transition to digital peer support, such as protecting confidentiality of service users [9]. Similar to other fields, supervision standards...
can assist in the development of knowledgeable and skilled
digital peer support specialists to clarify the role and
responsibility of digital peer support specialists and support
specialists in both an emotional and developmental capacity.
Introducing supervision standards for digital peer support has
the potential to help determine important administrative, educative, and supportive guidelines that supervisors could use
to maintain the practice of competent digital peer support. As
such, the aim of this study is to inform the development of
supervision standards for digital peer support and introduce
guidelines that supervisors can utilize to support, guide, and
develop competencies in digital peer support specialists.

Methods

Measures

Peer support specialists that currently offer digital peer support
services were recruited via email by an international listserv of
1500 peer support specialists, using a convenience sampling
method. Participants were eligible if they were (1) 18 years of
age or older and (2) a peer support specialist. Participants were
asked to complete a web-based presurvey with questions on
demographic information (eg, age, race, and gender) to ensure
variation in focus group participants and participated in a 1-hour
web-based focus group via a Health Insurance Portability and
Accountability Act (HIPAA)–compliant videoconferencing
platform in October 2020. The questions presented in the focus
groups were coproduced with 4 peer and nonpeer academic
scientists and 4 peer support specialists in a community-engaged
research approach described as Peer and Academic partnership
to help determine important administrative, educative, and
supportive guidelines supervisors could use to support the
practice of competent digital peer support and support digital
peer support specialists [10]. The interview guide contained
questions such as the following: “what essential knowledge
does a peer support specialist need to offer digital peer
support?” “what are the essential abilities peer support
specialists need to offer digital peer support?” “how do these
essential skills vary by lived experience (e.g., mental health,
physical health, substance use challenges, Veteran status, aging,
racial or ethnic diversity)?”

The collaborative development of supervision standards is
important within social-environmental historical contexts [11].
To reproduce such a group process and promote cross-individual
opinions, we used a series of focus groups to develop potential
supervision standards and guidelines. The analysis of digital
peer support supervision standards was based on four 1-hour
focus groups with a total of 59 participants, which took place
in October 2020. The focus group discussions followed Morgan’s
process model [12]. The focus groups were conducted by 2
authors using the interview guide. The interview guide was
successfully tested in a pretest. The focus group discussions
were recorded digitally, transcribed, and anonymized.
Researchers analyzed the data using Rapid and Rigorous
Qualitative Data Analysis (RADar), a team-based approach to
coding and analyzing qualitative data [13]. This approach was
selected because the RADar method produces qualitative results
thoroughly and quickly through its ability to organize, reduce,
and analyze data in user-friendly software packages such as
Excel (Microsoft Corp) [13]. The final set of supervision
standards is based on focus-group findings and used member
checking to ensure face validity and accuracy. Member checking
was used to increase the credibility of participant involvement
and data analysis [14]. Researchers presented data transcripts
to focus group participants for feedback, and participants were
asked to review the transcripts to determine if the researchers’
interpretation of the data match their intended meanings [14].

Analysis

Transcripts were formatted into an all-inclusive Excel sheet that
included column headings such as question, participant number,
and response. Team members assigned codes to each response.
After the all-inclusive Excel sheet was produced, the data table
was reduced to include only content relevant to the interview
questions. The remaining text and codes were then organized
into themes. In accordance with the RADar methodology,
themes were determined by the incidence of a code aligned with
an overarching theme (see Results). The process of member
checking was used to ensure the codes were interpreted correctly
and correctly organized into themes. Member checking is a
qualitative method used to validate findings, resolve conflicting
results, and assess the trustworthiness of qualitative results [15].
The percentage for each theme was found by dividing the
frequency in which the theme was present in the focus group
quotes by the total number of focus group quotes.

Ethical Considerations

The Committee for the Protection of Human Subjects at the
Dartmouth Health institutional review board approved the
project (STUDY02000620). Participants were told in the consent
form that they may voluntarily participate in the 1-hour
web-based focus group via a HIPAA-compliant
videoconferencing platform, and transcripts would be stored
via HIPAA-compliant software accessible only to the research
team.

Results

Participants

A total of 59 peer support specialists participated in the 4 focus
groups. Participants’ characteristics were reported prior to the
interview. The majority of participants were female (35/59,
76%), and the majority of participants had a minimum of high
school education (46/59, 100%). Participants were from 11
states and 3 countries.

Themes Covered

We identified 51 codes and a set of 11 themes related to the
development of supervision standards for digital peer support. Themes
covered administrative, educational, and supportive
functions that participants believed were integral to the
supervision of digital peer support specialists. The eleven themes
in order of most frequent to least frequent included: (1)
education on technology competency; (2) education on privacy,
security, and confidentiality in digital devices and platforms;
(3) education on peer support competencies and how they relate
to digital peer support; (4) administrative guidelines; (5)
education on the digital delivery of peer support; (6) education on technology access; (7) supervisor support of work-life balance; (8) emotional support (emerging); (9) administrative documentation (emerging); (10) education on suicide and crisis intervention (emerging); and (11) feedback (emerging).

Education on Technology Competency

The most prevalent theme was education on technology competency. This core theme constituted 21.8% (43/197) of the themes discussed in the focus groups. Peer support specialists believed it would be important for supervisors and agencies to help digital peer support specialists to develop knowledge on different technological platforms and devices. One peer support specialist mentioned:

A peer support specialist needs to have a comfort with whatever virtual platform they are using to interact with the service user. …I would imagine that, you know, the process of engagement would look very different depending on what kind of technology is being used as the shared platform.

Multiple participants added that digital peer support specialists should not only have basic knowledge of technology but should also be able to teach others how to use technology. One participant said:

They [digital peer support specialists] need some basic knowledge of computers and a virtual format, like accessing applications like zoom, not only to be able to navigate it, but to teach others how to access it and to navigate it on different platforms, like a computer and a mobile phone.

Access to training on digital technology and platforms was encouraged, and some peer support specialists suggested that agencies should offer videos and support in the acquisition of technology knowledge. One peer support specialist mentioned:

We offer different training videos and whatnot for different peer support levels…I think that’s something that would be interesting to see rolled out to other organizations as well. Having those sort of videos on how to use the technology exactly how to incorporate your thoughts and empathy into words and how to convey exactly what you’re hiding.

Education on Privacy, Security, and Confidentiality in Digital Devices and Platforms

The second theme, education on privacy, security, and confidentiality in digital devices and platforms, constituted 16.8% (33/197) of the themes discussed in the focus groups. Peer support specialists recommended that agencies and supervisors help to clarify and educate peers on certain topics related to privacy, security, and confidentiality. For example, a few participants suggested education on privacy policies, mandatory reporting, and data collection. One participant questioned:

There’s data, so how is data being collected, if it’s being shared? Are there third parties? Who’s being shared with? Is it shared with treatment teams? Is there data that need is like how does mandated reporting work for certain in certain circumstances with different agencies?…Is there a privacy policy?

Another participant recommended the clarification of definitions of privacy, security, and confidentiality and how they relate to the agency and to the service users in which peer support is being offered. One peer specialist said:

I think definitely, definitely just what the definition of confidentiality means and how it can be defined differently to different individuals…And I think having that mutuality between you what does confidentiality mean to you, what does it mean to me?

Developing knowledge in the evaluation of the security and privacy of digital platforms was not only important to peer support specialists but also having transparency around how privacy, security, and confidentiality guidelines may change depending on the form of peer support. One participant mentioned:

Differentiating between a warm line support call versus a crisis call…let people know this is a warm line. And we can talk about certain things. But as soon as you start to talk about harming yourself or others, now, this is turned into a crisis call. It’s no longer a warm line call. And what I’ve been told by emergency services workers in Virginia, one in particular, is that once there is a crisis, confidentiality, HIPAA no longer applies.

Education on Peer Support Competencies and How They Relate to Digital Peer Support

The third theme, education on peer support competencies and how they relate to digital peer support, constituted 12.7% (25/197) of the themes discussed in the focus groups. Many peer support specialists emphasized the importance of offering training on general peer support and maintaining the important values and principles of peer support during transitions to digital settings. One participant suggested:

Peer support requires many different skills. What essential knowledge do peer support specialists need to offer digital peer support? They need to know peer support cold. They need to be aware of how that translates virtually.

Other participants discussed the skills they would look for in peer support specialists when hiring them for peer support. They believed that a competent delivery of services would include general peer support competencies. One peer support specialist said:

I would want them to be top-notch peer supporter[s] before we even go into digital peer support. I would want someone that knows peers support, [and] that had the certified peer support specialist or a recovery coach so [he/she/one] knows how to deal with substance use, knows good peer support, and has a solid foundation. We’re not hiring people to do digital staff; we’re hiring people to do peer support and the digital platform is just one way to deliver the peer support.
While digital peer support requires new knowledge, skills, and supervisor support, it also requires solidification of the basic principles and values of peer support within the digital environment. One participant mentioned:

Peer support is peer support. I've been doing this for 23 years, I've talked to doctors, nurses, graduate students, et al and they all think that we are the saviors, and we’re going to be able to do the work that they can’t do. I’m just gonna sit down with them, I’m going to shut up, I’m going to listen, and I’m not going to fix them, I’m not going to judge them. So really, peer support, whether it’s digital or face to face, is the same basic principles.

Administrative Guidelines

The fourth theme was administrative guidelines. The theme of administrative guidelines constituted 10.7% (21/197) of the themes discussed in the focus groups. Peer support specialists recommended the development and transparency of administrative guidelines surrounding topics such as technology security, work-life boundaries, and suicide prevention. One participant said:

I think that’s essential knowledge for the peer specialist to know what they are, what boundaries they have, for what roles they may be having, and that they’re not all the same.

For example, participants suggested the creation of agency guidelines around communication methods used in suicide prevention on the internet. One participant mentioned, “having a real good fine communication flow chart. So that if an event does happen, that, you know, the peer support will know you know what to say, and also who to reach out to, so that they can get the extra support that they need through this as well” around confidentiality and what information peers need to share with supervisors. One peer support specialist said, “there are certain things that if we discussed and talk about I have to share with my supervisor” and around digital platforms and technologies that are acceptable for digital peer support. For example, one participant observed, “the state mandates what [digital platforms and technology] you can and can’t use when you’re providing services to their peers or to your peers.”

Education on the Digital Delivery of Peer Support

The fifth theme, education on the digital delivery of peer support, constituted 9.1% (18/197) of the themes discussed in the focus groups. With the transition from in-person peer support to digital peer support, peer support specialists have requested for agencies and supervisors to acknowledge the difference between providing services in-person opposed to digital. Many peer support specialists have appealed for training surrounding digital technologies and the digital delivery of peer support. For example, one participant mentioned:

I would want them to attend the digital peer support training. I think some training in digital support, the separate platforms and just talking about software and hardware and how they work together.

This includes training and skill development to engage service users on the internet and display empathy in a digital setting. One peer support specialist suggested:

Some sort of technical training on empathetic listening or some other ways of being able to convey emotion without necessarily being able to be connect. A personal conversation training maybe.

Participants also recommend that supervisors provide peer support specialists with the opportunity to decide whether digital peer support versus in-person peer support would be the best option for them, based on knowledge of both roles. One participant said:

Make sure any new hiring you’re having do digital support want to be a part of that. Not just ‘okay, we’ve hired you as a peer and you’re going to do this also’ because for some people it would be uncomfortable for them to be providing those services digitally.

Education on Technology Access

The sixth theme, education on technology access, constituted 8.6% (17/197) of the themes discussed in the focus groups. Access to technology and internet service is an important aspect of digital peer support. However, there are many populations that have difficulties accessing the devices and infrastructure they need to effectively use digital peer support. In order to expand the reach of digital peer support, peer support specialists believe peers need to be aware of the resources and tools available for the support of underserved populations. For example, one participant said:

it takes a really, really strong commitment and awareness to really open the doors of the service system much, much wider. Essentially, now that you know, there’s so much construction to the pandemic right.

Supervisors could potentially help digital peer support specialists to gain knowledge on resources and tools necessary to meet the needs of individuals, such as those experiencing homelessness or those living in rural settings. On participant mentioned:

There are a lot of individuals who may be on it, whether they’re experiencing homelessness, or don’t have consistent access to, for example, like charging or maybe have like, very, very basic phone, like flip phones. And so I would think to having kind of resourcefulness maybe being…having tools to be able to, I think, be proactive and kind of have insight and awareness to meet the needs of various populations in terms of location and in access.

Supervisor Support of Work-Life Balance

The seventh theme, supervisor support of work-life balance, constituted 8.6% (17/197) of the themes discussed in the focus groups. Supervisor support is essential to the maintenance of a healthy work-life balance. Many peer support specialists believe supervisors can help to set time limits and boundaries in service user’s access to telehealth. One participant recommended:
Incorporate, you know, some boundaries around your personal time, and to make it a main priority. And again, I think with the accessibility of the, of telehealth, it just is, you know, it just makes it just that much more important. ... the supervisor is going to have to help out with this.

Supervisors should be mindful of digital fatigue and help peer support specialists to schedule adequate breaks and time-limits. For example, one peer support specialist recommended:

...setting reasonable limits as to the amount of time that is expected of people. So giving people adequate breaks and stuff, so they don't have to do too much. ...just being mindful of when you are starting to get that fatigue.

Open communication and collaboration between supervisors and peer support specialists can help to address and resolve issues with work-life balance. One participant said:

...be upfront and open with those that you're supporting of your availability, because it's also important for you to protect yourself and your boundaries and maintain your health and self-care and all those things.

Emotional Support

The eighth emerging theme, emotional support, constituted 4.6% (9/197) of the themes discussed in the focus groups. Peer support specialists believe the additional challenges that come with digital peer support require additional support. Many participants believed reaching out for help from a supervisor or other peer can not only help with self-care but can help the peer support specialists to grow in their role as a digital peer support specialist. One participant said:

I think an essential ability is to that when you are having those tough moments to make sure that people reach out for help, so that they don't feel like they're struggling on their own, and that they do get feedback on some of those maybe tougher cases... don't be afraid to reach out for help, because it, it helps you grow in your role. But it also helps with the self-care.

Many participants agreed that the transition to digital peer support was often stressful and overwhelming. For example, one peer support specialist said:

Whether it's a digital or, you know, traditional, but I think more so digital, because a lot of us peer supports are new at this to, like, you know, it was hard enough for us to document now, and now we're trying to like, enter, you know, and do all this stuff on a phone or a computer, that is just a more stressful and I think it just makes me think of it, there's peers that have kind of, you know, resigned because of that, because it's like, Ah, you know.

As a result, many believed emotional supervisor support could help peer support specialists to address feelings such as stress and fatigue and other experiences such as retraumatization. One participant mentioned:

With peer support, you know, we talk about the trauma informed, because it's a part of our lived experience to share those pieces where it builds connection, we might get re-traumatized all over again, and not even realize it until we're trying to sleep that night that something doesn't feel right. You know, and so to be able to talk about it to reach out to your supervisor or another coach or peer, I think is really important and that's foundational.

Administrative Documentation

The ninth theme, administrative documentation, was an emerging theme and constituted 3% (6/197) of the themes discussed in the focus groups. Participants agreed that documentation of peer support would differentiate between in-person and web-based support. One participant mentioned, “I think documentation is going to be a little different for what you have to do on telehealth versus what you're doing in person.” Knowledge of the changes in documentation requirements and regulations is important to the supervisor role. For example, one participant suggested knowledge in “certain regulations...or documentation” and “knowing kind of what with your state are certain requirements or different kind of policies at the state level.” Participants suggested that digital peer support and the use of digital devices and platforms may offer challenges to the process of documentation for supervisors and others. One peer support specialist said, “It was hard enough for us to document and now we're trying…to do all this stuff on a phone or computer.”

Education on Suicide and Crisis Intervention

The tenth emerging theme, education on suicide and crisis intervention, constituted 2.5% (5/197) of the themes discussed in the focus groups. Many participants emphasized the difference between suicide and crisis intervention in-person opposed to digitally. For example, one participant said, “we're all very skilled in in-person crisis response and that completely changes when you're digital.” As a result, peer support specialists believe supervisors should offer trainings on digital suicide and crisis intervention. One participant recommended:

A training [or] even just a conversation about how to respond to someone that is in crisis, virtually. When you can't be there to control what's in their environment or their actions, how can you respond to keep them safe while you get them in person assistance? I think that's an important discussion that needs to be had and maybe training that needs to be developed.

Digital peer support specialists should know when to contact supervisors in an emergency. For example, one peer support specialist suggested the importance of “making sure that somebody is sufficient in those competencies such as suicide prevention, how to contact supervisors in an emergency, how to diffuse the situation, [and] talk somebody down” and have access to skill development and trainings in crisis intervention. One participant said:

One of the things that you would need to know is when to activate and do an active rescue versus just an
The purpose of this study was to inform the development of supervision standards endorsed by SAMHSA. The four focus groups reachedsaturation when additional data did not provide more information across international boundaries, and the themes identified could be used to promote consistency in the practice of digital peer support. Fifth, the data could not be stratified by volunteer-run services versus paid professional services or the role of the participant [16]. Sixth, recruitment occurred via a peer support specialist listserve—not solely service users of the mental health system. Including the voices of service users can enhance these competencies. Lastly, focus groups with specified digital peer support supervisors could help to expand the findings. Future research looking at the integration of digital peer support competencies within digital peer support supervision is needed. Future research should work to verify and build off of the digital peer support supervision standards and guidelines defined in this manuscript.

**Comparison With Prior Work**

Prior work has shown that digital support can be just as effective as in-person support in patient-clinician engagement [17]. There are, however, a few concerns about using the digital environment to facilitate health care. Digital health care may lead to social isolation without proper design in mHealth interventions [18]. Social isolation could be caused by the optional nature of digital peer support. Currently, supervision standards from SAMHSA for in-person peer support include the categories of administrative, educative, and supportive functions. However, the spread of digital peer support during the COVID-19 pandemic requires the expansion of supervision standards to include subthemes such as education on technology and privacy, support of work-life balance, and emotional support. Without digital supervision standards, there are potential risks of a breach in ethics and confidentiality, workforce stress, loss of productivity, loss of boundaries, and ineffectively serving users who participate in digital peer support services.

Digital peer support is quickly expanding across the globe. However, the transition to digital peer support brings new challenges and the necessary acquisition of new guidelines and skills. While SAMHSA has developed supervision guidelines for in-person peer support, digital peer support requires the expansion of supervision standards and the significance of administrative, educative, and supportive supervision. Digital peer support supervisors require specific knowledge and skills to communicate with service users and deliver peer support effectively. Therefore, supervisors also require new knowledge and skills to effectively develop, support, and manage the digital peer support role.

**Limitations**

This study is not without limitations. First, there are potentially other supervision standards and guidelines that have not been identified. Second, the sample lacked diversity based on racial and ethnic background. Future studies should consider the inclusion of disadvantaged populations such as Hispanic, and LGBTQIA populations and demographics such as homeless individuals. Third, as technology changes and digital peer support expands, supervision standards will need to be updated. Fourth, the findings cannot be generalized to all digital peer support specialists due to the small sample size. However, we reached saturation when additional data did not provide more information across international boundaries, and the themes identified could be used to promote consistency in the practice of digital peer support. Fifth, the data could not be stratified by volunteer-run services versus paid professional services or the role of the participant [16]. Sixth, recruitment occurred via a peer support specialist listserve—not solely service users of the mental health system. Including the voices of service users can enhance these competencies. Lastly, focus groups with specified digital peer support supervisors could help to expand the findings. Future research looking at the integration of digital peer support competencies within digital peer support supervision is needed. Future research should work to verify and build off of the digital peer support supervision standards and guidelines defined in this manuscript.

**Discussion**

**Principal Findings**

The following themes emerged from the four focus groups (N=59): (1) education on technology competency; (2) education on privacy, security, and confidentiality in digital devices and platforms; (3) education on peer support competencies and how they relate to digital peer support; (4) administrative guidelines; (5) education on the digital delivery of peer support; (6) education on technology access; (7) supervisor support of work-life balance; (8) emotional support (emerging); (9) administrative documentation (emerging); (10) education on suicide and crisis intervention (emerging); and (11) feedback (emerging). Established supervision standards may help to promote the competent delivery of digital peer support and help to encourage skill development, knowledge, support, and guidelines for the digital peer support role. These supervision recommendations may act to enhance the established supervision standards endorsed by SAMHSA.

Feedback also helps to uncover skills or knowledge supervisors themselves may need to improve upon. For example, 1 peer support specialist suggested that feedback “helps you as a supervisor to know what skills you need to work on both in supervision and you know for trainings.” Peer support specialists recommended feedback methods such as having a supervisor sit in on a meeting or call, or practicing role play in which the supervisor pretends to be in a crisis and then discusses the ways in which the specialist could have improved their digital support. One participant recommended having “the supervisor sit in on one or two or three of the peer support groups or calls” and another recommended having “a fake call or fake message conversation where for an hour the supervisor pretends to be in crisis and reaches out and we have to provide ample support to them and then they critique us on everything that we said after the hours up and they tell us if we are allowed to go on and do peer support provider.”

**Feedback**

The last emerging theme, feedback, constituted 1.5% (3/197) of the themes discussed in the focus groups. Many peer support specialists believe receiving feedback from supervisors is important to building competency in digital peer support specialists. One participant suggested supervisor feedback is helpful in “practicing how to do the things that you talk about. And making sure that somebody is sufficient in those competencies such as suicide prevention, how to contact supervisors in case there is a need for emergency, how to defuse the situation, how to talk somebody down although that’s the same thing with diffusing situation or what I’m saying and yeah that’s about it.”

The purpose of this study was to inform the development of supervision standards for digital peer support and introduce guidelines that supervisors can use to support, guide, and develop competencies in digital peer support specialists. Currently, supervision standards from SAMHSA for in-person peer support include the categories of administrative, educative, and supportive functions. However, the spread of digital peer support during the COVID-19 pandemic requires the expansion of supervision standards to include subthemes such as education on technology and privacy, support of work-life balance, and emotional support. Without digital supervision standards, there are potential risks of a breach in ethics and confidentiality, workforce stress, loss of productivity, loss of boundaries, and ineffectively serving users who participate in digital peer support services.

Digital peer support is quickly expanding across the globe. However, the transition to digital peer support brings new challenges and the necessary acquisition of new guidelines and skills. While SAMHSA has developed supervision guidelines for in-person peer support, digital peer support requires the expansion of supervision standards and the significance of administrative, educative, and supportive supervision. Digital peer support specialists require specific knowledge and skills to communicate with service users and deliver peer support effectively. Therefore, supervisors also require new knowledge and skills to effectively develop, support, and manage the digital peer support role.

**Discussion**

**Principal Findings**

The following themes emerged from the four focus groups (N=59): (1) education on technology competency; (2) education on privacy, security, and confidentiality in digital devices and platforms; (3) education on peer support competencies and how they relate to digital peer support; (4) administrative guidelines; (5) education on the digital delivery of peer support; (6) education on technology access; (7) supervisor support of work-life balance; (8) emotional support (emerging); (9) administrative documentation (emerging); (10) education on suicide and crisis intervention (emerging); and (11) feedback (emerging). Established supervision standards may help to promote the competent delivery of digital peer support and help to encourage skill development, knowledge, support, and guidelines for the digital peer support role. These supervision recommendations may act to enhance the established supervision standards endorsed by SAMHSA.

Feedback also helps to uncover skills or knowledge supervisors themselves may need to improve upon. For example, 1 peer support specialist suggested that feedback “helps you as a supervisor to know what skills you need to work on both in supervision and you know for trainings.” Peer support specialists recommended feedback methods such as having a supervisor sit in on a meeting or call, or practicing role play in which the supervisor pretends to be in a crisis and then discusses the ways in which the specialist could have improved their digital support. One participant recommended having “the supervisor sit in on one or two or three or four of the peer support groups or calls” and another recommended having “a fake call or fake message conversation where for an hour the supervisor pretends to be in crisis and reaches out and we have to provide ample support to them and then they critique us on everything that we said after the hours up and they tell us if we are allowed to go on and do peer support provider.”

**Feedback**

The last emerging theme, feedback, constituted 1.5% (3/197) of the themes discussed in the focus groups. Many peer support specialists believe receiving feedback from supervisors is important to building competency in digital peer support specialists. One participant suggested supervisor feedback is helpful in “practicing how to do the things that you talk about. And making sure that somebody is sufficient in those competencies such as suicide prevention, how to contact supervisors in case there is a need for emergency, how to defuse the situation, how to talk somebody down although that’s the same thing with diffusing situation or what I’m saying and yeah that’s about it.”

The purpose of this study was to inform the development of supervision standards for digital peer support and introduce guidelines that supervisors can use to support, guide, and develop competencies in digital peer support specialists. Currently, supervision standards from SAMHSA for in-person peer support include the categories of administrative, educative, and supportive functions. However, the spread of digital peer support during the COVID-19 pandemic requires the expansion of supervision standards to include subthemes such as education on technology and privacy, support of work-life balance, and emotional support. Without digital supervision standards, there are potential risks of a breach in ethics and confidentiality, workforce stress, loss of productivity, loss of boundaries, and ineffectively serving users who participate in digital peer support services.

Digital peer support is quickly expanding across the globe. However, the transition to digital peer support brings new challenges and the necessary acquisition of new guidelines and skills. While SAMHSA has developed supervision guidelines for in-person peer support, digital peer support requires the expansion of supervision standards and the significance of administrative, educative, and supportive supervision. Digital peer support specialists require specific knowledge and skills to communicate with service users and deliver peer support effectively. Therefore, supervisors also require new knowledge and skills to effectively develop, support, and manage the digital peer support role.

**Limitations**

This study is not without limitations. First, there are potentially other supervision standards and guidelines that have not been identified. Second, the sample lacked diversity based on racial and ethnic background. Future studies should consider the inclusion of disadvantaged populations such as Hispanic, and LGBTQIA populations and demographics such as homeless individuals. Third, as technology changes and digital peer support expands, supervision standards will need to be updated. Fourth, the findings cannot be generalized to all digital peer support specialists due to the small sample size. However, we reached saturation when additional data did not provide more information across international boundaries, and the themes identified could be used to promote consistency in the practice of digital peer support. Fifth, the data could not be stratified by volunteer-run services versus paid professional services or the role of the participant [16]. Sixth, recruitment occurred via a peer support specialist listserve—not solely service users of the mental health system. Including the voices of service users can enhance these competencies. Lastly, focus groups with specified digital peer support supervisors could help to expand the findings. Future research looking at the integration of digital peer support competencies within digital peer support supervision is needed. Future research should work to verify and build off of the digital peer support supervision standards and guidelines defined in this manuscript.

**Comparison With Prior Work**

Prior work has shown that digital support can be just as effective as in-person support in patient-clinician engagement [17]. There are, however, a few concerns about using the digital environment to facilitate health care. Digital health care may lead to social isolation without proper design in mHealth interventions [18]. Social isolation could be caused by the optional nature of digital peer support.
interfacing with others when using technology and inability to connect using visual insights and nonverbal cues [19]. Prior work has expressed the importance of training health care workers on demonstrating digital empathy to address differences from in-person interaction [19]. Additional security must also be implemented in mHealth interventions including HIPAA compliance in videoconferencing software and digital patient records, to ensure privacy and confidentiality [20]. Therefore, digital peer support may increase the capacity for engagement with individuals while providing quality relationships and satisfactory care if supervision standards are improved to accommodate digital security and address the potential for concerns such as social isolation and empathy [5]. In addition to improving quality of care, proper supervision may facilitate peer recovery specialist practice. In recent peer recovery specialist literature, it was expressed in focus groups that consistent supervision that emphasizes self-care and principles learned in training may lead to greater worker retention and job satisfaction [21].

Conclusions
Introducing supervision standards for digital peer support is a first step in helping to guide the delivery of digital peer support and the development of digital peer support specialists. As defined by SAMHSA, supervision is important to competency building and skill development in peer support. Supervision has the potential to improve performance, empower workers, and promote knowledge of the peer role. The shift to digital peer support has expanded the reach of in-person support and has shifted how peer support is both delivered and managed. The identification of supervision guidelines for digital peer support has the potential to facilitate the transition from in-person to digital peer support and promote best practices in both digital peer support delivery and the supervision of digital peer support specialists.

Conflicts of Interest
None declared.

References


Abbreviations

HIPAA: Health Insurance Portability and Accountability Act
RADar: Rapid and Rigorous Qualitative Data Analysis
SAMHSA: Substance Abuse and Mental Health Services Administration

©Caroline Collins-Pisano, Michael Johnson, George Mois, Jessica Brooks, Amanda Myers, Deanna Mazina, Marianne Storm, Maggie Wright, Nancy Berger, Ann Kasper, Anthony Fox, Sandi MacDonald, Sarah Schultze, Andrew Bohm, Julia Hill, Karen Fortuna. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 19.06.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Perceived Patient Workload and Its Impact on Outcomes During New Cancer Patient Visits: Analysis of a Convenience Sample

Safa Elkefi, MSc; Onur Asan, PhD
School of Systems and Enterprises, Stevens Institute of Technology, Hoboken, NJ, United States

Corresponding Author:
Onur Asan, PhD
School of Systems and Enterprises
Stevens Institute of Technology
1 Castle Point Terrace
Hoboken, NJ, 07030
United States
Phone: 1 4145264330
Email: oasan@stevens.edu

Abstract

Background: Studies exploring the workload in health care focus on the doctors’ perspectives. The ecology of the health care environment is critical and different for doctors and patients.

Objective: In this study, we explore the patient workload among newly diagnosed patients with cancer during their first visit and its impact on the patient’s perceptions of the quality of care (their trust in their doctors, their satisfaction with the care visits, their perception of technology use).

Methods: We collected data from the Hackensack Meridian Health, John Theurer Cancer Center between February 2021 and May 2022. The technology use considered during the visit is related to doctors’ use of electronic health records. A total of 135 participants were included in the study. Most participants were 50-64 years old (n=91, 67.41%). A majority (n=81, 60%) of them were White, and only (n=16, 11.85%) went to graduate schools.

Results: The findings captured the significant effect of overall workload on trust in doctors and perception of health IT use within the visits. On the other hand, the overall workload did not impact patients’ satisfaction during the visit. A total of 80% (n=108) of patients experienced an overall high level of workload. Despite almost 55% (n=75) of them experiencing a high mental load, 71.1% (n=96) reported low levels of effort, 89% (n=120) experienced low time pressure, 85.2% (n=115) experienced low frustration levels, and 69.6% (n=94) experienced low physical activity. The more overall workload patients felt, the less they trusted their doctors (odds ratio [OR] 0.059, 95% CI 0.001-2.34; P=.007). Low trust was also associated with the demanding mental tasks in the visits (OR 0.055, 95% CI 0.002-2.64; P<.001), the physical load (OR 0.194, 95% CI 0.004-4.23; P<.001), the time load (OR 0.183, 95% CI 0.02-2.35; P=.046) the effort needed to cope with the environment (OR 0.163, 95% CI 0.05-1.69; P<.001), and the frustration levels (OR 0.323, 95% CI 0.04-2.55; P=.03). The patients’ perceptions of electronic health record use during the visit were negatively impacted by the overall workload experienced by the patients (OR 0.315, 95% CI 0.08-6.35; P=.01) and the high frustration level experienced (OR 0.111, 95% CI 0.015-3.75; P<.001).

Conclusions: The study’s findings established pathways for future research and have implications for cancer patients’ workload. Better technology design and use can minimize perceived workload, which might contribute to the trust relationship between doctors and patients in this critical environment. Future human factors work needs to explore the workload and driving factors in longitudinal studies and assess whether these workloads might contribute to unintended patient outcomes and medical errors.

(JMIR Hum Factors 2023;10:e49490) doi:10.2196/49490

KEYWORDS
health care; cancer patients’ workload; trust; satisfaction; health information technology
Introduction

Background
Cancer is a major global public health issue in modern medicine [1]. Based on a report by the National Cancer Institute, 18.1 million new cancer cases were recorded in 2018, with 9.5 million cancer-related deaths worldwide [2]. This number is expected to rise to more than 20 million new cancer cases by 2025 [3] and 29.5 million by 2040 [2]. After initial diagnosis, clinical information becomes complex, leading to increasingly complicated treatment recommendations for patients with cancer [4]. The ecology of the first visits after diagnosis is unique [5] since patients experience significant life disruptions [6]. These disruptions can come from disease symptoms and the burden of treatment-related decision-making [6]. In these new cases, a diagnosis threatens their physical well-being and their sense of cognitive and emotional well-being [1]. In addition, they have difficulty understanding the medical information and generally report dissatisfaction with the delays in prognosis and follow-ups [7]. This results in psychosocial concerns among patients [1,8]; they experience high distress, emotional stress, uncertainty about mortality, and a disturbing social life [9,10]. These cognitive and emotional workloads might overburden patients with cancer, resulting in a higher likelihood of nonadherence to treatment plans [11].

Within the context of cancer care, the link between people, work, and goals is complex and multidimensional. Studying how humans interact with their environment, including the tools, technology, and systems they use, is referred to as human factors. Human factors are critical in understanding the interactions between health care personnel, patients, and the broader health care system in cancer care [12]. For example, according to human factors research, effective communication and teamwork among health care workers are critical for obtaining optimal patient outcomes in cancer care settings [12]. Furthermore, creating clear goals and addressing cancer patient needs and preferences during the visit is critical for increasing patient engagement and outcomes. Human factors study aids in the identification of potential hurdles and challenges in the cancer care process, such as workload, information overload, and other issues [12]. By addressing these issues, health care institutions can increase patient safety, reduce medical errors, and improve overall cancer treatment quality [12].

Cancer visits involve 3 main parties: doctors delivering information, patients, and families receiving the services under emotionally pressured situations, and technology supporting the information delivery and overall care. The primary interaction occurs between the doctor and patient, discussing the new diagnosis and future treatment plan. Electronic health records (EHRs) are the main technologies used by doctors during the visit. However, some studies reported that EHR use might increase doctors’ cognitive workload [13], negatively impact doctor-patient communication [14], and create less attentive doctors during the visit. Studies also showed that oncologist doctors use EHRs less than primary care doctors during these emotional visits to avoid the aforementioned negative aspects [15].

To deliver optimal holistic cancer care, it remains essential to take actions centered around the patients, mirroring their needs and expectations [16]. Patient-centered care is based on respect for patients’ expectations and values. It aims to provide them with the needed education and information, ensure their continuous secure access to care, and involve their families to support their emotional well-being [17]. In cancer care, the relationship between doctors and patients discriminates between 2 underlying dimensions: technical, related to the medical situation, and affective, pertaining to the relations and emotions of the patients [18]. Thus, the rational-consumer patient-centered care model would not suit oncology settings [19]. Patient-centered care has proved to be important in improving health care outcomes. When doctors engage in effective communication and shared decision-making and demonstrate trust in their patients, patients show more efficacy in self-management and have better psychological and physical health outcomes [20-23]. Patient-centered care should also be studied from a patient ergonomics perspective. Patient ergonomics is the application of human factors or related disciplines to study and improve patients’ and other nonprofessionals’ performance of effortful work activities in pursuit of health goals [24,25]. A central emerging concept of societal views of health care considers that the patients actively perform “work” to achieve health-related goals and objectives [26]. By that, human factors position the patients in the center of the work system aiming to improve their experience with the load of work assigned [24,27]. In highly sensitive situations like cancer care, this paradigm can help us better understand the dynamics between the 3 actors of the visits (doctor, patient, technology) and how their interaction can influence critical outcomes like quality of care, trust of doctors, and acceptability or perception of technology use.

Advancements in digital communication and medical technologies have led to digitalizing health care [14,28]. With the increased adoption and use rate of EHRs in cancer care, oncologists can use the provided data in the critical decision-making process and support their workload [29]. In a study by Mazur et al [30], the enhancement of EHR systems’ usability was associated with better oncologist doctors’ cognitive workload and performance. Studies also explore how EHR influences doctors’ cognitive workload and performance in various settings [31]. However, no study has explored patients’ overall workload as well as how technology use impacts their workload during the visits. Given the importance of supporting new cancer patients’ “work” success, a holistic approach that recognizes the impact of workload on care outcomes in the first visits remains important. Therefore, this cross-sectional survey-based study investigates the workload of cancer patients in new cancer patient visits and its association with the following outcomes: trust in care doctors, satisfaction with the care delivered, and their perception of the technology (EHR) used in cancer care.

Theory and Hypotheses
It is critical to understand the users’ workload while performing a task using technology, especially in highly complex environments such as health care. The purpose of a workload evaluation is to determine the user’s workload while he or she
is working on a given task using or utilizing a system or technology [32]. The concept of workload has been described as “the cost of performing a task in this way that reduces the capacity to perform other tasks that use the same processing resource” [32]. The workload is measured to assess the performance of users and systems [33]. Since working memory is limited, distractions, new information, and complex information can interfere with clinical decision-making and can result in errors [34]. Cognitive load is a measure of how many cognitive resources are used during thinking, learning, problem-solving, and reasoning [35]. Studies used subjective workload assessments such as NASA TLX (National Aeronautics and Space Administration Task Load Index) in various contexts, including aviation and health care [36,37]. In health care, most studies focus on measuring clinician workload [38]. However, there is a lack of studies focusing on understanding patients’ perspectives of workload. Especially no study measured patients’ workload in high-anxiety environments such as cancer care [39].

Problems related to workload-related vulnerabilities are discussed in cancer care literature [40]. Discovering a cancer diagnosis brings emotional pressure to new patients and causes a stress load that makes them experience difficulty finding their emotional stability [41]. In addition, trust in doctors is an important component of patient-centered care as it plays a pivotal role in the success of cancer treatment strategies [42]. In this study, we hypothesize that high levels of workload during the initial visit would negatively impact newly diagnosed cancer patients’ trust in their doctors on the first visit after diagnosis (hypothesis 1).

Furthermore, as a new cancer diagnosis is disorienting for patients, newly diagnosed patients might experience high levels of anxiety and depression [43]. With the triggered unmet physical, psychological, and informational needs, patients require much more attention than what they receive [44]. In addition, new patients report dissatisfaction with care systems (delays in diagnosis, follow-ups, etc) driven by confusing, unclear processes and inefficient procedures [7,45]. We hypothesize that satisfaction with the care visit is negatively impacted by the workload experiences of newly diagnosed cancer patients in the very first visits after diagnosis (hypothesis 2).

Finally, we showed in a previous review that health information technology is used in cancer care to propose solutions that can strengthen the cancer patients’ relationship with their doctors, empower their well-being and build a structured target-oriented care process for them [46]. Despite its benefits, using EHR extensively during these highly emotional visits might have negative consequences. Newly diagnosed cancer patients’ experienced physical, mental, and emotional pressure can affect their perceptions towards using technologies like EHRs during the visits. Thus, we hypothesize that newly diagnosed cancer patients’ high workload negatively impacts their perception of EHR use during the very first visits after diagnosis (hypothesis 3).

To sum up, the 3 hypotheses of this study investigate the interrelation between the 3 actors of the visit: new cancer patients, doctors, and technology. Figure 1 details the conceptual framework followed.

Figure 1. Overall conceptual framework of the study.

Methods

This study took place at the Hackensack Meridian Health, John Theurer Cancer Center. The setup of the patients’ rooms in the cancer center is standard and identical to each other with an EHR system in the room.

Ethics Approval

The study obtained ethical approval from both the Stevens Institute of Technology and the Hackensack Meridian John Theurer Cancer Center IRB offices (IRB ID 00011536).

Data Collection and Participants

This study is part of a federal grant that was funded in May 2020. Due to COVID-19 restrictions, the data collection started with 7 months delay in February 2021. We used a convenience sampling method to recruit newly diagnosed patients when they came for their very first visit with a cancer doctor at the Cancer Center. Convenience sampling is a nonprobability sampling technique that involves selecting your research sample based on convenience and accessibility [45]. The inclusion and exclusion criteria included (1) having the new cancer patient...
visit, (2) understanding English, (3) being between 18-65 years old, and (4) not having any dementia and cognitive impairments. Patients who have upcoming visits are first contacted by phone and informed about the study. If they agreed, they completed a consent form to participate in the study and completed the survey within 24 hours of their first visit. We strictly used 24 hours rule to capture their initial experience fresh right after their very first visit with their cancer doctor. Due to COVID-19 restrictions, we have administered the survey over the phone. Each participant completing the survey was given a US $30 gift card. Data collection was conducted from February 2021 through May 2022. No participant identifiers were obtained during the study. Based on Green’s rule of thumb, for regression and correlation analysis, the sample size should be larger than 50 participants [47]. In our study, we aimed for 130 to 150 participants. By May 2022, we had received 135 participants. The participants were seen by 13 doctors. We limited the number of patients seen by each doctor to a maximum of 15 patients per doctor. We recruited patients with various cancer diagnoses. However, the majority of them were diagnosed with breast cancer, lymphoma, and multiple myeloma. We had 58 female participants (Table 1) and 45 participants from minority groups (Hispanic and African American). Most participants were between the ages of 50-64 years old.

Table 1. The demographics of the participants included in the study.

<table>
<thead>
<tr>
<th>Demographics</th>
<th>Participants (N=135), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
</tr>
<tr>
<td>18-34</td>
<td>7 (5.19)</td>
</tr>
<tr>
<td>35-49</td>
<td>35 (25.93)</td>
</tr>
<tr>
<td>50-64</td>
<td>91 (67.41)</td>
</tr>
<tr>
<td>&gt;64</td>
<td>2 (1.48)</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
</tr>
<tr>
<td>No diploma</td>
<td>4 (2.96)</td>
</tr>
<tr>
<td>Some school</td>
<td>17 (12.59)</td>
</tr>
<tr>
<td>High school</td>
<td>44 (32.59)</td>
</tr>
<tr>
<td>Technical college</td>
<td>20 (14.81)</td>
</tr>
<tr>
<td>Bachelor</td>
<td>34 (25.19)</td>
</tr>
<tr>
<td>Grad school or more</td>
<td>16 (11.85)</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
</tr>
<tr>
<td>Black American</td>
<td>28 (20.74)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>17 (12.59)</td>
</tr>
<tr>
<td>White</td>
<td>81 (60)</td>
</tr>
<tr>
<td>Other</td>
<td>9 (6.67)</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>77 (57.04)</td>
</tr>
<tr>
<td>Female</td>
<td>58 (42.96)</td>
</tr>
</tbody>
</table>

**Instrumentation**

We developed our survey using validated instruments from the literature. The questions included in this survey measure the perceived workload, trust towards doctors, EHR use perception, and patient satisfaction with the care received. We also captured the participants’ demographics (education level, age, race, and gender).

The perceived workload is captured through the NASA TLX index. NASA’s TLX index is a popular construct in human factors science [48]. It was shown to be among the most reliable and valid questionnaires to measure workload in health care settings [49]. As shown in Table 2, the NASA TLX index has 6 main components physical demand, temporal demand, mental demand, effort, frustration, and performance. Trust is captured through the doctors’ trust scale, and the technology used is captured through the perception of the computer use scale. The exact questions used to capture each variable are detailed in Multimedia Appendix 1.
### Table 2. Variables of the study.

<table>
<thead>
<tr>
<th>Category and variable</th>
<th>Scale or questions used</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Workload</strong></td>
<td>NASA TLX(^a) index</td>
</tr>
<tr>
<td>Physical demand</td>
<td></td>
</tr>
<tr>
<td>Temporal demand</td>
<td></td>
</tr>
<tr>
<td>Mental demand</td>
<td></td>
</tr>
<tr>
<td>Effort</td>
<td></td>
</tr>
<tr>
<td>Frustration</td>
<td></td>
</tr>
<tr>
<td>Performance</td>
<td></td>
</tr>
<tr>
<td><strong>Quality of care</strong></td>
<td></td>
</tr>
<tr>
<td>Doctor’s trust</td>
<td>Trust scale</td>
</tr>
<tr>
<td>Technology use perception</td>
<td>EHR(^b) use a perception scale</td>
</tr>
<tr>
<td>Satisfaction with care</td>
<td>How satisfied were you with the overall visit?</td>
</tr>
</tbody>
</table>

\(^a\) NASA TLX: National Aeronautics and Space Administration Task Load Index.  
\(^b\) EHR: electronic health record.

We adopted NASA TLX to capture workload experience by measuring mental, physical, temporal, performance, effort, and frustration components [50]. The NASA TLX has been validated for single-task environments [50,51]. The questions of the NASA scale compose an averaged 100 point-score. Originally, researchers applied a weighting procedure to the raw test scores of NASA TLX to develop a composite score tailored to individual workload definitions, however many researchers have eliminated the weighting procedure and instead use the raw test scores since it is simpler to apply: the ratings are averaged or added to create an estimate of overall workload between 0-100 [49]. In addition, we dichotomized the variables as follows: a value of 30 points and more is considered a high workload [52]. We also followed the same logic for the cut-off of high and low for specific components of NASA TLX. Trust in doctors is measured in this study using the subscale “trust in health care providers” of the “Multidimensional Trust in Health Care Systems Scale,” developed and validated by Egede and Ellis [53]. It is an averaged score composed of 10 questions with 4 Likert scale answers [53]. We dichotomized the trust scale in a way that a score above 50% was considered a high trust. Technology use perception is measured through the averaged scale of “Patient-Reported Satisfaction with Physician Computer Use,” assessed and validated for electronic medical records and other computer uses in health care settings to evaluate patients’ perception of doctors’ use of computer systems [54]. For satisfaction with care, we use a 5-Likert scale question where patients are asked about their satisfaction with the visit. Both satisfaction with care and technology perception scales are dichotomized in a way that a score above 50% is considered high. We test the overall score and the components’ associations for each variable. Figure 2 shows the detailed conceptual framework of the study.
The Nature of the First Cancer Visit and Tasks

It is essential to understand the nature of the visit and tasks in the first cancer visit to envision the workload for the patients. The first consultation with new cancer patients is spent on the following tasks:

1. Reviewing diagnosis of cancer, type of cancer, extent of cancer
2. Reviewing imaging studies performed and discussing any additional work-up that might be recommended (eg, breast magnetic resonance imaging, additional biopsies, other imaging studies)
3. Discuss treatment options (surgery, radiation, systemic treatment, plastic surgery), assuming most of the work-up is completed.
4. Assess general health status or other medical issues
5. Assess social support or mental health or coping
6. Assess for any clinical trials

The primary task for patients is engagement during these tasks. Some of these tasks are done by shared decision-making, so patients are required to understand discussed topics for their best interests.

Statistical Analysis

First, we ran descriptive statistics for all the study variables. Second, logistic regression analysis was run for the scores and the components to explore the correlation between all the variables and test the hypothesis as shown in the framework (Figure 2). All the regression models were adjusted for the demographics (age, race, gender, and education level). Model variables were dichotomized for analysis purposes based on the information existing in the literature [55]. Confirmatory factor analysis (CFA) was performed using the survey measures to analyze the psychometric properties of the variables. The fit and reliability of the CFA to the data were determined as acceptable as indicated by commonly used metrics such as composite reliability greater than 0.90 [56], average variance extracted greater than 0.50 [57], Guttman lambda 6, and coefficient omega (for second-order CFA of expectancy) greater than 0.80 [58]. All data cleaning and analyses were done using Python 3.7 using some packages (eg, pandas, stats, numpy).

Results

Descriptive Analysis

Figure 3 shows the distribution of the overall workload across the participants. The lowest workload we observed was around 20-25 out of 100 (7/135, around 5% of the participants), whereas the highest level of workload was around 65-70 out of 100. Overall, the majority of patients reported a high workload (score >30).
Table 3 shows the percentages of participants who have low and high workloads across different demographics. As shown in Table 3, (108/135, 80%) of patients experienced an overall high level of workload based on NASA TLX scores. However, when we look at the specific components, we see that 55% (75/135) of the patients experienced a high mental load, which is the question of mental activity to perform activities such as thinking, deciding, remembering, etc. On the other hand, we also see that participants reported a low level of effort (71.1%, 96/135); time pressure felt due to the rate or pace at which tasks occurred during the visit (88.89%, 120/135); low frustration due to feeling insecure, discouraged, irritated, stressed, and annoyed (115/135, 85.2%); and the perception that low levels of physical activity were required from them to perform activities in the visit (94/135, 69.93%).
Table 3. The distribution of the workload trends among the different demographic subgroups.

<table>
<thead>
<tr>
<th>Demographics</th>
<th>NASA TLX&lt;sup&gt;a&lt;/sup&gt; score, n (%)</th>
<th>Frustration, n (%)</th>
<th>Performance, n (%)</th>
<th>Effort, n (%)</th>
<th>Time load, n (%)</th>
<th>Physical load, n (%)</th>
<th>Mental load, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low (100)</td>
<td>High (100)</td>
<td>Low (100)</td>
<td>Low (100)</td>
<td>Low (100)</td>
<td>Low (100)</td>
<td>Low (100)</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-34 (n=7)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>5 (71.43)</td>
<td>2 (28.57)</td>
<td>85.71 (14.29)</td>
<td>57.14 (42.86)</td>
</tr>
<tr>
<td>35-49 (n=35)</td>
<td>9 (25.71)</td>
<td>26 (74.29)</td>
<td>31 (88.57)</td>
<td>40 (11.43)</td>
<td>35 (71.43)</td>
<td>28.57 (28.57)</td>
<td>57.14 (42.86)</td>
</tr>
<tr>
<td>50-64 (n=91)</td>
<td>18 (19.78)</td>
<td>73 (80.22)</td>
<td>74 (82.42)</td>
<td>17 (17.58)</td>
<td>2 (2.20)</td>
<td>97.80 (25.7)</td>
<td>77.14 (22.86)</td>
</tr>
<tr>
<td>&gt;64 (n=2)</td>
<td>0 (0)</td>
<td>2 (100)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black American (n=28)</td>
<td>6 (21.42)</td>
<td>22 (78.57)</td>
<td>26 (92.59)</td>
<td>2 (7.41)</td>
<td>27 (96.30)</td>
<td>88.89 (11.11)</td>
<td>79.26 (20.74)</td>
</tr>
<tr>
<td>Hispanic (n=17)</td>
<td>4 (23.53)</td>
<td>13 (76.47)</td>
<td>14 (88.24)</td>
<td>3 (11.76)</td>
<td>17 (100)</td>
<td>64.71 (35.29)</td>
<td>76.47 (23.53)</td>
</tr>
<tr>
<td>White (n=81)</td>
<td>14 (17.95)</td>
<td>67 (82.05)</td>
<td>68 (83.33)</td>
<td>13 (16.67)</td>
<td>1 (1.28)</td>
<td>98.72 (28.21)</td>
<td>64.10 (35.90)</td>
</tr>
<tr>
<td>Other (n=9)</td>
<td>2 (22.22)</td>
<td>7 (77.78)</td>
<td>6 (66.67)</td>
<td>3 (33.33)</td>
<td>9 (100)</td>
<td>66.67 (33.33)</td>
<td>66.67 (33.33)</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (n=77)</td>
<td>19 (24.68)</td>
<td>56 (75.32)</td>
<td>58 (85.71)</td>
<td>11 (14.29)</td>
<td>0 (0)</td>
<td>66.23 (33.77)</td>
<td>55.73 (44.16)</td>
</tr>
<tr>
<td>Female (n=58)</td>
<td>8 (13.79)</td>
<td>50 (86.21)</td>
<td>49 (84.48)</td>
<td>9 (15.52)</td>
<td>2 (3.45)</td>
<td>96.55 (6.55)</td>
<td>68.97 (31.03)</td>
</tr>
<tr>
<td>All (N=135)</td>
<td>27 (20.80)</td>
<td>108 (85.19)</td>
<td>115 (14.81)</td>
<td>20 (14.88)</td>
<td>2 (14.88)</td>
<td>98.52 (71.11)</td>
<td>69.63 (30.37)</td>
</tr>
</tbody>
</table>

<sup>a</sup>NASA TLX: National Aeronautics and Space Administration Task Load Index.

Impact of Workload on Quality of Care

Figure 4 shows the results of the different models we tested. The first model shows the relationship between the overall NASA TLX score and its relationship with 3 outcomes. The other 3 models show the relationship between each component of NASA TLX (mental load, physical load, time load, effort, performance, and frustration) and outcome measures (trust, satisfaction, and perception of technology use). As shown in Figure 4, the more overall workload patients felt, the less they trusted their doctors (odds ratio [OR] 0.059, 95% CI 0.001-2.34; \(P=0.007\)). We, thus, fail to reject hypothesis 1. Low trust was also associated with the demanding mental tasks in the visits (OR 0.055, 95% CI 0.002-2.64; \(P<0.001\)), the physical load (OR 0.194, 95% CI 0.004-4.23; \(P<0.001\)), the time load (OR 0.183, 95% CI 0.02-2.35; \(P=0.046\)), the effort needed to cope with the environment (OR 0.163, 95% CI 0.05-1.69; \(P<0.001\)), and the frustration levels (OR 0.323, 95% CI 0.04-2.55; \(P=0.03\)).
Patient’s performance during the visits did not impact their trust in their doctors.

**Figure 4.** Detailed results for the conceptual model’s validation. HIT: health information technology. *P*<.05; *P*<.01; *P*<.001.

When we look at the model for satisfaction, the overall workload level did not impact the patients’ satisfaction with the overall visit. We, thus, reject hypothesis 2. The detailed satisfaction model also showed that only 1 specific component significantly impacts care satisfaction. The patient’s satisfaction with the overall visit was negatively impacted by the time load they experienced (OR 0.123, 95% CI 0.001-2.56; *P*=.04), as shown in [Figure 4](#).

Finally, patients’ perceptions of EHR use during the visit were negatively impacted by the overall workload experienced by the patients (OR 0.315, 95% CI 0.08-6.35; *P*=.01) and the high frustration level experienced (OR 0.111, 95% CI 0.015-3.75; *P*<.001), as shown in [Figure 4](#). We, thus, fail to reject hypothesis 3.

**Discussion**

**Principal Results**

Doctor workload has been studied by several studies in the human factors field. However, there is a clear gap in the literature investigating the workload of patients during visits, especially in a complex environment such as cancer. This is the first study to explore cancer patients’ workload and its associations with various outcomes (doctors’ trust, use of technology perception, satisfaction with care) during the visit.

**Summary of Findings**

Encounters in cancer care might be stressful and cognitively highly demanding for patients and doctors. Studies have already shown that doctors have moderate to high workloads, even in primary care settings [38]. In our study, we also observed that most of the participants (108/135, 80%) experienced a somewhat high workload during the visit across various demographics.

The various models we tested yielded interesting results. The overall NASA TLX workload scores had a significant association with the patient’s trust in doctors as well as the patient’s perception of technology use (doctor’s EHR use) during the visit. However, we did not observe a significant association with the satisfaction score.

According to our findings, the high workload perceived by patients during the visit results in less trust in their doctors. The detailed components of NASA TLX, including patients’ frustration in addition to the effort, mental, physical, and time load required to perform activities during the visits, also impacted the patients’ trust in their doctors. This interesting finding has implications for reconsidering and redesigning the structure of the first visits. Building trust and rapport between doctor and patient on the first visit is critical and requires high-quality communication skills [59]. In addition, many factors were shown to impact trust in the literature due to its fragility, such as the rapid changes in the health care system and conditions of care [60]. In response to the cancer diagnosis, patients experience emotional and physical impairment coupled with developing a sense of transitoriness (finitude of life) [61]. It becomes hard for them to adapt to the new situation and find continuity in their lives in the middle of the flow of information and decisions they should deal with [61]. This may explain the association between the high workload and low trust noticed among new cancer patients. A study by Plomp and Ballast [62] investigating the vulnerability of doctor-patient trust in occupational health showed that in critical sensitive situations, a high workload creates a vulnerability in patients, resulting in more difficulty trusting their doctors. The authors state that “a combination of poor health and high workload could create a greater (need to) trust but is obviously not a sufficient condition to overcome stubborn distrust” [62].
We also noted a significant relationship between workload and patients’ perception of the use of technology during the visits. In fact, new cancer patients experience a range of emotions, including shock, sadness, anger, disappointment, and confusion [63]. The added anxiety of not knowing the next steps can cause even more stress and frustration [63]. The emotional burden was found to be highly associated with their perception of the quality of care and life among newly diagnosed lung cancer patients [63]. As new cancer patients would still be building their communication paths with their doctors, technology use during the visit might add to the high workload and improve the frustration of the patients during these emotional visits. This also might indicate that patients may not prefer technology used within the visits to be able to spend more time with their doctors and feel well listened to. In addition, the detailed model also yielded an interesting result showing only 1 component of NASA TLX: the frustration variable concerning if the patients felt annoyed, stressed, or discouraged, which has a significant association with perceived technology use. This is an interesting result supporting some of the early studies done in primary care. Despite the potential role of technology in strengthening the therapeutic alliance between doctors and their patients [46], researchers have argued that using computers during visits, especially under emotional situations, may negatively impact interaction as it does not allow the patients to find their way of decoding nonverbal information appropriately and may prevent them from building cue channels of interactions with their doctors [64].

Finally, the high level of workload did not impact patient satisfaction with the visits. Only the time load negatively impacts satisfaction. This also shows that time pressure during cancer visits might influence satisfaction negatively. Given that this is their first visit as cancer patients, they want to use all necessary time to discuss their concerns and do not want to feel rushed during the visit. Some studies also argued that cancer patients’ satisfaction with care is associated with the timeliness of care, as cancer patients have a load that exceeds the time available to them [65]. In addition to the increased susceptibility to stress resulting from the diagnosis, the patient’s anxiety can be amplified by long waiting times for appointments and results and long medical visits, which negatively impacts the patients’ satisfaction with the quality of care delivered [66].

Even though the NASA TLX index was designed specifically for aviation occupations, it has proved its use in different industries [67,68]. In healthcare, it was shown to be effective in measuring doctors’ workload in various critical environments to explore the impact of technology use on their activities [69]. In a study by Lund et al [70], it was used to measure the workload levels of surgeons to evaluate the association between their burnout and their performance. It showed high levels of workload after long working shifts. It was used by Norasi et al [71] to evaluate the usability of the robots to support the surgeons’ workload and teamwork effectiveness. It was also used to test the effectiveness of using augmented reality technologies to support cognitive demand [72]. Thus, in addition to its role in evaluating the usability of technology in health care, we showed that the NASA TLX index has the potential to support researchers in evaluating the workload of patients in cancer care.

Practical Implications

Theoretically, it is feasible to presume that newly diagnosed cancer patients experience a high workload. However, in practice, it remains important to investigate the impact of the high workload on patients’ quality of care perception to suggest corrective strategies based on the patients’ needs and performance. Our findings also have theoretical implications. First, most of the studies investigating workload in health care explore it from a doctor’s perspective accounting for their performance boosters to create a good work environment. Our study is the first study in the field of human factors that investigates workload among patients and captures its direct impact on their perception of care quality (trust in doctors, satisfaction with care, perception of technology use). Identifying the direct factors impacted by workload adds to the literature on the predictors of the quality of cancer care. Learning what influences the overall rating of care can enable doctors to accommodate vulnerable patient groups. Identifying health care aspects that are independently associated with the overall rating of care may enable targeted efforts when planning and prioritizing initiatives to improve the patient-experienced quality of care. Furthermore, as technology use was associated with a high workload in our analysis, more thought should be given to better design simplification and better system integration to control the physical and cognitive workload among patients as well as doctors. The clear impactful interactions between doctors, patients, and technology raise a flag for the importance of considering this trio in the different interventions made in cancer care to make sure to involve all parts of the equation. This will make “patient work” less demanding and more accurate, which includes understanding the situation and making the right shared decision in the cancer treatment during the first cancer visit.

Limitations and Future Studies

This study has some limitations that should be acknowledged. First, the study is cross-sectional and captures the patients’ opinions at a certain point in time. Future studies should involve longitudinal data and explore the proposed relationships over time to compare the same findings throughout different stages of cancer (treatment vs diagnosis) and observe the evolution. Second, patients participated in the study at a very early stage after diagnosis. Despite the originality of the findings, this may add more bias to their perception of their workload. A follow-up after some days should be done to validate their perceptions. Some environmental factors, like the crisis related to COVID-19, may add more pressure to the patient’s situation, which may bias the results related to the emotional load and the frustration level. Better control of environmental factors would increase the validity of the data from various measurements. Apart from addressing our limitations, there is room for additional future research based on our findings. Future research also should explore the workload of doctor and patient dyads who are on the same visit to compare the workload assessment and factors leading to workload in both parties. Researchers should also test various technology designs and explore how their use might...
improve the perceived workload of both doctors and patients during the visits.

Conclusions

We showed that most patients with cancer in the study experienced a high workload based on NASA TLX scores. The overall workload is also associated significantly with patient trust in the doctor as well as the perception of EHR use during the visit, but it does not impact satisfaction significantly. Future human factors work might explore the workload and driving factors in longitudinal studies and assess whether these workloads might contribute to unintended patient outcomes and medical errors. Finally, better technology design and use can minimize perceived workload, which might contribute to the trust relationship between doctors and patients in this critical environment.

Acknowledgments

Research reported in this publication is supported by the National Institute of Nursing Research of the National Institutes of Health under award 1R15NR018965-01. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We also thank Dr Scott Rowley, Mariska Kop, Alene Mathurin, and Trina Mcclendon for their support in the data collection process.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey questions.

References


47. Van Voorhis CW, Morgan BL. Understanding power and rules of thumb for determining sample sizes. TQMP 2007;3(2):43-50 [FREE Full text] [doi: 10.20982/tqmp.03.2.p043]


Accepted version, https://humanfactors.jmir.org/2023/1/e49490

©Safa Elkefi, Onur Asan. Originally published in JMIR Human Factors (https://humanfactors.jmir.org). 18.08.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Mental Health Professionals’ Attitudes Toward Digital Mental Health Apps and Implications for Adoption in Portugal: Mixed Methods Study

Diogo Nogueira-Leite1,2,3*, MSc; José Miguel Diniz1,2,3*, MSc, MD; Ricardo Cruz-Correia1,4, MSc, PhD

1Department of Community Medicine, Information and Decision in Health, Faculty of Medicine, University of Porto, Porto, Portugal
2Nova School of Business and Economics Health Economics and Management Knowledge Center, New University of Lisbon, Lisbon, Portugal
3Programme in Health Data Science, Faculty of Medicine, University of Porto, Porto, Portugal
4Center for Health Technology and Services Research, Porto, Portugal
*these authors contributed equally

Corresponding Author:
Diogo Nogueira-Leite, MSc
Department of Community Medicine, Information and Decision in Health
Faculty of Medicine, University of Porto
Rua Dr. Plácido da Costa
Porto, 4200-450
Portugal
Phone: 351 225513600
Email: diogo.fdnl@gmail.com

Abstract

Background: Digital health apps are among the most visible facets of the ongoing digital transition in health care, with mental health–focused apps as one of the main therapeutic areas. However, concerns regarding their scientific robustness drove regulators to establish evaluation procedures, with Germany’s Digitale Gesundheitsanwendungen program pioneering in app prescription with costs covered by statutory health insurance. Portugal gathers a set of conditions and requirements that position it as an excellent test bed for digital health apps. Its daunting mental health landscape reinforces the potential interest in new interventions. To understand if they would be acceptable, we need to understand the supply side’s attitudes and perceptions toward them, that is, those of psychiatrists and psychologists.

Objective: This study aims to understand the attitudes and expectations of psychiatrists and psychologists toward digital mental health apps (DMHAs) in the Portuguese context, as well as perceived benefits, barriers, and actions to support their adoption.

Methods: We conducted a 2-stage sequential mixed methods study. Stage 1 consisted of a cross-sectional web survey adapted to the Portuguese context that was delivered to mental health professionals and psychologists. Stage 2 complemented the insights of the web survey results with a key opinion leader analysis.

Results: A total of 160 complete survey responses were recorded, most of which were from psychologists. This study aims to understand the attitudes and expectations of psychiatrists and psychologists toward digital mental health apps (DMHAs) in the Portuguese context, as well as perceived benefits, barriers, and actions to support their adoption.

Objective: This study aims to understand the attitudes and expectations of psychiatrists and psychologists toward digital mental health apps (DMHAs) in the Portuguese context, as well as perceived benefits, barriers, and actions to support their adoption.

Methods: We conducted a 2-stage sequential mixed methods study. Stage 1 consisted of a cross-sectional web survey adapted to the Portuguese context that was delivered to mental health professionals and psychologists. Stage 2 complemented the insights of the web survey results with a key opinion leader analysis.

Results: A total of 160 complete survey responses were recorded, most of which were from psychologists. This is the most extensive study on mental health professionals’ attitudes and perceptions of DMHAs in Portugal. A total of 87.2% (136/156) of the respondents supported the opportunity to prescribe DMHAs. Increased health literacy (139/160, 86.9%), wider adherence to treatment (137/160, 85.6%), and proper disease management (127/160, 79.4%) were the most frequently agreed upon benefits of DMHAs. However, only less than half (68/156, 43.6%) of the respondents planned to prescribe or recommend DMHAs, with psychologists being more favorable than psychiatrists. Professionals faced substantial barriers, such as a lack of information on DMHAs (154/160, 96.3%), the level of initial training effort (115/160, 71.9%), and the need for adjustments of clinical processes and records (113/160, 70.6%). Professionals reported that having more information on the available apps and their suitability for health objectives (151/160, 94.4%), more scientific evidence of the validity of the apps as a health intervention (147/160, 91.9%), and established recommendations of apps by specific clinical guidelines or professional societies (145/160, 90.6%) would be essential to foster adoption.

Conclusions: More information about DMHAs regarding their clinical validity and how they work is necessary so that such an intervention can be adopted in Portugal. Recommendations from professional and scientific societies, as well as from governmental bodies, are strongly encouraged. Although the benefits of and the barriers to using these apps are consensual, more evidence, along with further promotion of mental health professionals’ digital literacy, is needed.
Introduction

Digital Health Apps

Digital health apps substantiate, perhaps more tangibly than most solutions developed so far, the opportunities the digital age may bring for human health [1-3]. Chief among them is the potential to make health care simultaneously more accessible and personalized. The magnitude of the business-to-consumer market speaks for itself; according to IQVIA’s Digital Health Trends 2021 report [4], >350,000 health apps are available in various app stores, with 110 apps downloaded >10 million times and accounting for approximately 50% of all downloads.

The same report [4] highlights that the COVID-19 pandemic gave a solid impulse to app use, with mental health, cardiovascular disease, and diabetes management–related apps accounting for approximately 50% of disease-focused offerings in app stores. The number of downloads and the variety of apps allowed many to conclude that there is an apparent demand at the consumer level. Simultaneously, patients’ perspectives on app use in settings where prescriptions are available [5,6] reflect that patient acceptance still has a long way to go, although some evidence on how to enhance it is already available.

Regulatory Approaches

The realization that most tools in the field needed to be more mature to match formulated expectations prompted specialists to increasingly voice concerns that most apps need to meet more clinical and technical validation standards, often lacking any empirical support for their adoption [7-11]. A growing consensus is that regulation is needed, especially for apps that diagnose, treat, or manage high-risk conditions [12,13]. Governments and regulators have started to define policy frameworks to determine the benefits of digital solutions [14-18], trying to understand ways to reduce uncertainty around digital health app use and kick-start discussions on their potential payment or reimbursement.

Germany took the lead by implementing a regulatory framework specific to digital health apps and their market access and reimbursement. Its Digitale Gesundheitsanwendungen (DiGA) program [19] was inaugurated in October 2020, and on January 23, 2023, a total of 40 apps qualified for statutory insurance reimbursements. Of these 40 apps, 18 (45%) were classified as DiGA for mental disorders [20]. France is studying a replication of the DiGA approach and has a preliminary reimbursement process through its assessments of medicotechnical and medical benefits [21,22].

Belgium ranks second in implementation; although mHealthBelgium [23] was launched in 2018, it officially started conducting appraisal and reimbursement processes in January 2021, with its selection process based on a 3-level validation pyramid [24]. Most European countries have so far opted for softer, more decentralized approaches, with legal obligations and compliance rules based on the General Data Protection Regulation [25] or the Medical Devices Regulation [26]. By contrast, Singapore and the United States resort to their medical device regulations. The Food and Drug Administration has been particularly active, basing its assessment of apps and digital therapeutics on the Software as a Medical Device framework [27,28]. In September 2022, it updated its Policy for Device Software Functions and Mobile Medical Applications [29] and its guidance on clinical decision support software [30], divulged its key findings from the precertification pilot program at the federal level [31], and launched its Digital Health Policy Navigator for developers [32].

Promise of Digital Mental Health Apps to Aid Care Delivery

Mental disorders are one of the areas where the penetration of digital health apps is most prevalent [4,20]. Reasons for their apparent popularity range from the stigma of seeking treatment and individual privacy needs to the convenience of doing it from everywhere and the diversity of treatments available (eg, meditation, cognitive behavioral therapy, group therapy, teleconsultation, etc) [33-38]. These disorders are also one of the disease areas in desperate need for increased and enhanced access. This need already existed before the COVID-19 pandemic, and many have pointed to the deleterious impact of the pandemic on mental health as one of its considerable long-term consequences [39-41]. The burden of disease it entails, both before and after the COVID-19 pandemic, and the way it impacts many other health conditions make it a priority for action [42-44].

Portugal is often cited as a country where mental disorders, particularly anxiety and depression, are above average; the prevalence of mental health disorders in 2019 was estimated at 8.27% of disability-adjusted life years and 19.27% of disease cases. The statistics for anxiety and depressive disorders were expected to be 2.58% and 3.16% of the total disability-adjusted life years and 9.08% and 5.88% of the disease prevalence, respectively [45]. A summary of its comparison with the global, European Union (EU), and German landscape is presented in Table 1.

Conversely, there is limited access to psychological and psychiatric care, with waiting times ranging from 13 to 237 days for a psychiatry consultation in the Portuguese National Health Service from July to September 2022 [46]. The time frame for a psychology consultation in the 11 institutions that reported it for the same period ranged from 15 to 134 days. Considering most depression and anxiety cases, albeit...
responsible for most of the disease prevalence of mental disorders, are classified as nonpriority cases, waiting times can be expected to range from 44 to 237 days.

In a country burdened by out-of-pocket payments about double the EU average [47,48], the possibility of resorting to private sector providers is minimal, as insurers cover only some associated costs. It is necessary and urgent to find new solutions. The combination of disease prevalence and lack of access to care, along with a relatively digitized health system and average indicators of digital literacy, makes Portugal an excellent test bed to understand whether digital mental health apps (DMHAs) can, or cannot, help people receive the care they need.

One of the key promises of digital health apps is increased access. However, no innovative intervention in health—be it a drug, medical device, or any other—achieves critical mass without the endorsement of health professionals [49-51]. Therefore, it becomes essential to understand, from the perspective of mental health professionals (here defined as psychologists and psychiatrists), their level of comfort with digital health apps, their main challenges in adopting them, and what can enable and enhance their use.

To our knowledge, only one study has been performed on the Portuguese landscape of web-based interventions for psychologists [52]. No studies were found concerning the attitudes of psychiatrists in Portugal or combining the attitudes and perspectives of Portuguese psychiatrists and psychologists toward DMHA as a specific web-based intervention. Our study aimed not only at bridging these gaps but also at contributing importantly to do so (1) after the COVID-19 pandemic and its catalyzing effect on telehealth adoption [53]; (2) after major prescription and reimbursement processes were enacted in the EU space; and (3) by mapping the supply side of web-based mental health care, given the mediating effect of mental health professionals [50]. Our study contributes substantially to researchers, academia, industry, and policy makers by providing necessary information on how to leverage the DMHA as a tool to increase access to mental health care and improve patient outcomes while reducing the burden of disease associated with mental health disorders.

### Objective

This paper aimed to understand mental health professionals’ attitudes (defined as psychologists and psychiatrists in this study) toward DMHAs in the Portuguese context. Mental health professionals will be questioned regarding perceived benefits, barriers to adoption, and potential ways of supporting the adoption of DMHAs. These apps were the focus of this study. The authors aimed to achieve this by directly inquiring mental health professionals regarding their specific clinical practices, perceived needs, and expectations.

### Methods

#### Study Design

The research team used a mixed methods methodology. Stage 1 consisted of a cross-sectional web-based survey adapted to the Portuguese context and delivered to mental health professionals and psychologists. It used a web-based quantitative data–focused survey, adapted to the Portuguese context, which served as a primary data source. Stage 2 used the answers collected from the survey to help conduct a qualitative key opinion leader (KOL) analysis.

As per the research protocol [54], the methods initially intended for this study had to be adapted because of the survey’s low response rate. Both the web-based survey and the structure of the KOL analysis followed the same constructs studied by Dahlhausen et al [55] to maximize comparability with that study and the German landscape, albeit focused on mental health.

Notably, this study did not include a literature review of technology adoption, relevant case studies, or subsequent interviews with mental health professionals and psychologists on their views and perceptions toward DMHAs. This was deemed appropriate, as such processes had the objective of building up the questionnaire, and we intended to apply a translated version of the survey to the Portuguese context. Following the original publication, we conducted a web-based survey on a pretest group of health care professionals.

In Portugal, psychiatrists and psychologists are expected to prescribe or otherwise interact with DMHAs and act upon the patients’ mental health. To maximize the targeting of these professionals, we restricted our approach to these 2 groups of health care workers. Moreover, to complement our interpretation of the survey results and help us understand the meaning and generalizability to the national context, we conducted a KOL analysis with a select set of professionals belonging to 1 of the 2 surveyed groups, with roles in clinical practice, academia, industry, or a combination of these.

Furthermore, given that no prescription processes are established in Portugal for DMHAs, it is not possible to rigorously define who would be authorized to recommend or prescribe DMHAs. Therefore, we asked clinicians to answer questions that report

---

**Table 1.** Share of disability-adjusted life years (DALY) and disease prevalence (in percentage points) per condition and geography. Data source: Institute for Health Metrics and Evaluation (IHME). GBD Compare Data Visualization. Seattle, WA: IHME, University of Washington, 2020.

<table>
<thead>
<tr>
<th></th>
<th>Portugal</th>
<th>Germany</th>
<th></th>
<th>EU†</th>
<th>World</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DALY (%)</td>
<td>Prevalence (%)</td>
<td>DALY (%)</td>
<td>Prevalence (%)</td>
<td>DALY (%)</td>
</tr>
<tr>
<td>Mental disorders</td>
<td>8.27</td>
<td>19.27</td>
<td>6.43</td>
<td>15.59</td>
<td>6.65</td>
</tr>
<tr>
<td>Depression</td>
<td>3.16</td>
<td>5.88</td>
<td>2.16</td>
<td>4.32</td>
<td>2.42</td>
</tr>
<tr>
<td>Anxiety</td>
<td>2.58</td>
<td>9.08</td>
<td>1.95</td>
<td>7.07</td>
<td>1.69</td>
</tr>
</tbody>
</table>

†EU: European Union.
to recommendation or prescription according to their own cases, as psychiatrists are allowed to prescribe medication in Portugal, whereas psychologists are not. Our results should be interpreted accordingly.

Moreover, in our survey, we did not ask about health insurance coverage status, as it proves more relevant, in the Portuguese context, to understand whether they work for the National Health Service, in private practice, or both. As previously stated, we targeted only psychiatrists and psychologists for this survey, with the latter comprising most of the respondents (127/158, 80.4% of the answers). Although no data regarding the number of psychologists are available at the time of this study’s conclusion, it is our perception and that of the KOLs that the largest share of mental health professionals would be attributable to this group of practitioners.

In addition, we chose to represent survey data differently, intending to highlight the distribution of the categorical (Likert scale) answers and define their centrality without recurring to arithmetic operations.

**Web-Based Survey Design**

The first part of the study comprised a cross-sectional, web-based survey. We used the final survey questionnaire available in the Multimedia Appendix 1 in the study by Dahlhausen et al [55] as given and translated it to Portuguese using a licensed translator (Multimedia Appendix 2).

This translation was delivered to 10 mental health professionals—5 from each professional group, psychologists and psychiatrists—to gather their input. Mental health professionals’ feedback was focused on calibrating the survey to (1) reflect essential questions to ask regarding the use of digital health tools by mental health professionals and (2) adapt to a Portuguese mental health care context. This allowed us to focus solely on mental health and DMHAs. The survey used by Dahlhausen et al [55] depicted, although implicitly and more pragmatically, the theoretical constructs of the Unified Theory of Acceptance and Use of Technology [56]. Given that our adaptation process did not affect this, we considered our questionnaire, by the same token, to adapt to the same theory and its constructs. The obtained feedback was incorporated to produce a final survey questionnaire for this study, available in English in Multimedia Appendix 1. Therefore, several changes were made, including modifying and adding questions, per the survey reviewers’ suggestions. Although these limit the direct comparability between studies, they reflect the different needs and issues of the 2 countries. Both translations—the questionnaire by Dahlhausen et al [55] to Portuguese for adaptation and the final adapted survey questionnaire in Portuguese to English—were performed by SPS Traduções, a specialized translation firm.

Before broad diffusion, the survey questionnaire was pretested by 5 different colleagues to determine the completion time and identify shortcomings. As a result, an introductory page on digital health apps and developments in their regulatory landscape was included to provide initial baseline information before the start of the survey. The completion time was estimated to be between 4 and 7 minutes. To establish a basis for comparison with a reimbursable app system, mental health professionals were asked to consider a scenario in which these apps fulfilled regulatory requirements and addressed safety, quality, and efficacy concerns. Accordingly, mental health professionals’ responses are to be interpreted under this assumption and not necessarily to these apps’ current form as available in Portugal. Nonetheless, it could be argued that because both Portugal and Germany belong to the EU and its internal market, an app developer would want to maximize comparability between apps, tweaking them for populational specificities.

Several web-based channels and methods were used to distribute the questionnaire to health care professionals. The survey’s link was circulated in the newsletter of the Portuguese Order of Psychologists and through the social media of several members of the Psychiatry Specialty College of the Portuguese Order of Medical Doctors. The professionals who engaged in the questionnaire’s adaptation were invited to perform snowball recruiting by sharing the survey link through their social media accounts and with professional contacts and forums where they were involved.

In addition, Knok healthcare [57], a fully integrated telemedicine platform company, offered to disseminate the questionnaire on its social media accounts to its relevant audience of health care professionals. This free initiative is part of Knok’s mission to deliver social impact by divulging the potential benefits of telemedicine. Finally, the Portuguese Society of Psychiatry and Mental Health agreed to disseminate the survey via social media on Twitter.

The platform used was Inqueritos@UP, the University of Porto’s internal survey manager by LimeSurvey. The survey adhered to and was reported following the Checklist for Reporting Results of Internet E-Survey (CHERRIES) guidelines. The period for answer collection ran from September 26, 2022, to November 6, 2022, the same 6-week period applied in the study by Dahlhausen et al [55].

The study’s Data Protection and Privacy Policy, made available in the Ethics Approval, Informed Consent, and Participation section, comprises all relevant information on these aspects. To maximize responses, the only inclusion criterion was to be registered with the mental health professional’s respective professional order. No exclusion criteria were introduced, and no financial incentives were offered. Figure 1 summarizes the survey’s adaptation and communication workflow.

The gathered data were analyzed according to the methods used in the study by Dahlhausen et al [55] to allow for maximum comparability between the results. Descriptive statistical analyses were performed for all variables, whenever possible. Estimates of association for the variables corresponding to “Results” subsections in the study by Dahlhausen et al [55] were also computed.

Only data excluded because of different health system organizations and their consequences for mental health professionals (eg, statutory health insurance in Germany vs little to no point-of-care payments in Portugal) or of reasonable
suggestion during the feedback period were treated differently and according to the nature of each variable.

The correspondence map between the initial questionnaire (ie, by Dahlhausen et al [55]) and the final survey questionnaire is presented in the table in Multimedia Appendix 3 [54].

Furthermore, data were analyzed to find associations among variables, especially between health care professionals’ demographic and professional characteristics, attitudes toward DMHAs, and the likelihood of prescription. These were conducted on RStudio (version 2022.07.1 build 554; RStudio Inc) using chi-square tests or, when conditions for using chi-square tests were not met, Fisher exact tests with Monte Carlo approximation and 2000 replicates [58,59]. R packages used for data processing, analysis, and graphical representations were tidyverse, data.table, png, gt, gtExtras, gtssummary, Hmisc, likert, grid, forcats, scales, reshape2, and rcompanion.

KOL Analysis

The KOL analysis [60] served two purposes: (1) to compile what the prominent opinion voices in psychiatry and psychology in Portugal and across academia, clinical practice, and industry understand to be the main benefits, adoption barriers, and measures that can support the adoption of DMHAs and (2) to gather their input on the conducted survey’s results to understand whether they agree with their perception of most Portuguese psychiatrists’ and psychologists’ views on DMHAs.

The KOL analysis followed a 2-step approach. The first step consisted of semistructured individual web-based interviews, followed by a second round of confirmation of the gathered consensus. The method used for the KOL analysis could be defined as a compromise between the Nominal Group Technique and the Delphi Technique [61].

In the first part of this interview, each KOL was asked what were, in their opinion, the top 3 benefits, barriers to adoption, and measures that could support the adoption of DMHAs in the Portuguese context. In the second part of the interview, the interviewers showed KOLs the results of the conducted web survey and asked them to comment on the results. The requested comments were focused on whether these results agreed with their perception of most Portuguese psychiatrists’ and psychologists’ views on DMHAs.

The second stage of the KOL analysis consisted of circulating the main elements gathered during performed interviews and asking for their comments in free text, namely, whether they agreed with the established consensus and whether something important was missing. This analysis was divided into benefits, barriers to adoption, and support measures to ensure methodological consistency.

A total of 25 KOLs were identified and selected to participate in this research (practicing psychiatrists or psychologists, researchers, and managers with a psychiatry or psychology background working in digital health companies). Invitations to participate were made via email through the identification of publicly available professional email addresses. Snowball recruiting was used to find more participants; every contacted KOL was asked to suggest other KOLs that could be reached for this study. KOLs were given a 7-day period to answer whether they wanted to participate and, if so, to point to 3 dates and times for the interview. Those who did not respond to the initial invitation received a follow-up email after 3 days to increase the response rate. Interviews ran from November 25 to December 23, 2022, lasting between 35 and 70 minutes.

Overall, 4 of the KOLs were psychiatrists and 2 were psychologists. This is an important feature to remember, as the number of psychiatrists who answered the survey was much smaller than that of the psychologists. This allowed us to expand the interpretation power of the answers provided by psychiatrists. Notably, every KOL contributes regularly to the public discourse on mental health and the use of technology to tackle problems around mental ill health, for example, in written media. Moreover, half of the KOLs (3/6, 50%) have already developed DMHAs or more comprehensive digital health tools, and 5 (83%) out of 6 KOLs work in academia and private or public sector.

Ethics Approval, Informed Consent, and Participation

The Ethics Committee of the Faculty of Medicine of the University of Porto pronounced itself favorable to the research project on June 30, 2022 (Opinion 52/CEFMUP/2022).
Ethical considerations and safeguards for the study and its supporting documents (including the web-based survey) were encoded in the study’s Data Protection and Privacy Policy, which received approval from the Data Protection Officer of the University of Porto and is transcribed as follows:

- To preserve participants’ privacy, they will not be asked to provide any personally identifiable information. In addition, participants will not be tracked for having started or completed the survey, increasing privacy but limiting the possibility of reminders.

- **Informed consent and consenting capacity:** all potential participants (mental health professionals and academic community members) will be given web-based written information on the study and its objectives and will be asked to provide consent (click to agree) that they are willing to participate, do so freely and voluntarily. Nonparticipation will not compromise their current roles. Participation in the study will be voluntary, and no inducements or incentives to participate will be offered.

- **Confidentiality:** Any data or personal details that could potentially reveal the identity of individuals will be removed. Only anonymized, deidentified information will leave the place of origin. A database with responses will be maintained on a password-protected database. All research data will be stored on a password-protected desktop computer at the host organization. Study participants will be invited, through a link provided on the last page of the survey, to provide their name and electronic address to allow the research team to facilitate their receipt of a synopsis of the study findings on publication. This list will be kept separately on a password-protected database and a password-protected desktop computer at the host organization. All data will be stored securely at the host institution and destroyed 3 years after the PhD defense date. It is estimated that the PhD will be defended between October 2023 and December 2023.

- **General Data Protection Regulation compliance will be adhered to in terms of the following:**
  
  - Data privacy rights: participants will have the right to request information about their data throughout the research process.
  
  - Transfer of data: participants will be informed about the circumstances under which their data may be transferred and safety measures that will be taken to protect the data (eg, data are encoded).
  
  - Retention of data: Participants will be informed of the duration for which their data will be stored.

Using Inquéritos@UP, survey data were stored at the university’s servers and thus not shared with external entities, constituting another layer of privacy protection. Furthermore, the survey’s first page briefly explained the required data and the rationale behind it.

### Results

#### Web-Based Survey

**Demographics**

A total of 160 health care professionals completed the questionnaire, with only some nonresponses to specific questions. Although the overall survey response rate could not be determined, given its means of distribution and the adopted privacy-ensuring settings, 400 people opened the survey. This translates into a completion rate of 40%, making this study the most extensive on mental health care professionals’ attitudes and expectations toward DMHAs in Portugal.

Table 2 shows the characteristics of those who completed the questionnaire and their work. The most common age group was the 36-45 years segment (59/160, 36.9%), closely followed by the 26-35 years segment (56/160, 35%), being skewed toward a younger population. Of the 160 participants, 134 (83.7%) participants were female, likely presenting a higher representation in the sample than in the national presentation (52.8% of all psychiatrists in Portugal were female, with no publicly available data for psychologists) [62]. Most respondents (136/160, 85%) served populations with >20,000 residents, representing a primarily urban setting.

Many practitioners worked in >1 type of practice, most commonly at clinics (57/160, 35.6%) and hospitals (56/160, 35%), with only 10.6% (17/160) working at the primary care level. A considerable portion of professionals were involved with private consultation services, either individually (35/160, 21.9%) or in a group (39/160, 24.4%). However, in the study by Dahlhausen et al [55], most clinicians were split between single or joint practice environments. Among those who answered about the number of mental health professionals and psychologists they worked with, most (64/119, 53.8%) reported having ≥5 such professionals in their workplace, with 26.9% (32/119) of participants reporting >10 professionals in their workplace. Of these 160 participants, 127 (79.4%) were psychologists and 25 (15.6%) were medical psychiatrists. The distribution of answers was relatively homogeneous with regard to the number of mental health professionals in the workplace. In the replicated paper, approximately half of the respondents (613/1268, 48.3%) reported having only 1 practitioner.
Table 2. Demographic and professional characteristics of the sample (N=160).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td></td>
</tr>
<tr>
<td>&lt;26</td>
<td>4 (2.5)</td>
</tr>
<tr>
<td>26-35</td>
<td>56 (35)</td>
</tr>
<tr>
<td>36-45</td>
<td>59 (36.9)</td>
</tr>
<tr>
<td>46-55</td>
<td>31 (19.4)</td>
</tr>
<tr>
<td>56-65</td>
<td>6 (3.8)</td>
</tr>
<tr>
<td>&gt;65</td>
<td>4 (2.5)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>134 (83.8)</td>
</tr>
<tr>
<td>Male</td>
<td>26 (16.2)</td>
</tr>
<tr>
<td>Size of population covered (inhabitants)</td>
<td></td>
</tr>
<tr>
<td>&lt;5000</td>
<td>5 (3.5)</td>
</tr>
<tr>
<td>5001-20,000</td>
<td>0 (0)</td>
</tr>
<tr>
<td>20,001-100,000</td>
<td>36 (25.5)</td>
</tr>
<tr>
<td>100,001-500,000</td>
<td>52 (36.9)</td>
</tr>
<tr>
<td>&gt;500,000</td>
<td>48 (34)</td>
</tr>
<tr>
<td>Unknown</td>
<td>19 (11.9)</td>
</tr>
<tr>
<td>Workplace</td>
<td></td>
</tr>
<tr>
<td>Hospital</td>
<td>56 (35)</td>
</tr>
<tr>
<td>Primary care</td>
<td>17 (10.6)</td>
</tr>
<tr>
<td>Clinic</td>
<td>57 (35.6)</td>
</tr>
<tr>
<td>Individual private consultation</td>
<td>35 (21.9)</td>
</tr>
<tr>
<td>Collective private consultation</td>
<td>39 (24.4)</td>
</tr>
<tr>
<td>Number of doctors and psychologists in workplace</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>19 (16)</td>
</tr>
<tr>
<td>2</td>
<td>15 (12.6)</td>
</tr>
<tr>
<td>3</td>
<td>13 (10.9)</td>
</tr>
<tr>
<td>4</td>
<td>8 (6.7)</td>
</tr>
<tr>
<td>5</td>
<td>10 (8.4)</td>
</tr>
<tr>
<td>6</td>
<td>6 (5)</td>
</tr>
<tr>
<td>7</td>
<td>5 (4.2)</td>
</tr>
<tr>
<td>8</td>
<td>4 (3.4)</td>
</tr>
<tr>
<td>9</td>
<td>4 (3.4)</td>
</tr>
<tr>
<td>10</td>
<td>3 (2.5)</td>
</tr>
<tr>
<td>&gt;10</td>
<td>32 (26.9)</td>
</tr>
<tr>
<td>Unknown</td>
<td>41 (25.6)</td>
</tr>
<tr>
<td>Profession</td>
<td></td>
</tr>
<tr>
<td>Psychologist</td>
<td>127 (79.4)</td>
</tr>
<tr>
<td>Psychiatrist</td>
<td>25 (15.6)</td>
</tr>
<tr>
<td>Other</td>
<td>6 (3.8)</td>
</tr>
<tr>
<td>Unknown</td>
<td>2 (1.3)</td>
</tr>
</tbody>
</table>
**Perceived Potential Benefits From DMHAs and Attitudes Toward DMHAs**

Figure 2, as it happens to the following figures (ie, Figures 3-5), provides a visual representation of the Likert scale responses, ordered according to the color-coded legend. For each question (eg, increased health literacy), an overall share is represented for negative (“Totally Disagree” or “Disagree”), neutral (“Don’t Know” or “Neither Agree nor Disagree”), or positive (“Agree” or “Totally Agree”) answers. Their percentage is displayed on the left, central, and right positions of the stacked bars, respectively. As far as central tendency measures are concerned, we selected the median value of each group of answers besides the corresponding bar. The neutral response group was set as the center of the axis to facilitate comparisons between answers, with distribution graphs representing the distribution of the provided answers.

Potential benefits to the patient from DMHAs, namely, improved ability to make informed choices, proper disease management, improved treatment adherence, improved access to health care, and increased health literacy were perceived very positively by responding health professionals. This was demonstrated by the overall positive perceptions toward using DMHA services; in no case did the general agreement have <65% of the responses. A higher general agreement proportion of the answers (including “Agree” and “Totally Agree”) was found concerning the gains in health literacy (139/160, 86.8%) and treatment adherence (137/160, 85.6%).

With regard to general disagreement ratios (including “Disagree” and “Totally Disagree”), the most unfavorable perceptions (24/160, 15%) were demonstrated toward the capacitation of informed choices by the patients. This question (32/160, 20%) and another regarding improved access to health care (29/160, 18.1%) represented the highest proportions of neutral answers (including “Don’t Know” and “Neither Agree nor Disagree”).

For the general disagreement ratios, the highest proportion was described concerning the possibility of improving the quality of patient assistance (27/160, 16.9%), closely followed by the previously mentioned improved treatment success. The share of neutral answers ranged from 21.3% (34/160; regarding better time management owing to efficiency gains) to 46.3% (74/160; regarding improved treatment success).

Our adapted survey did not assess the perceptions on additional new patients, or additional income for this group of benefits, having added others regarding the satisfaction of higher DMHA-based demand and the DMHA as an additional treatment option at the suggestion of our survey reviewers.
Figure 3. Perceptions of potential benefits for health care professionals.

Figure 4. Perceived barriers to prescription.

Figure 5. Measures to support adoption. DMHA: digital mental health app.

Prescription Intentions of DMHAs

Of the 160 participants, 68 health care professionals (n=62, 91% psychologists; n=6, 9% psychiatrists) declared to have an increased likelihood (“Likely” or “Very Likely”) to prescribe DMHAs in the coming 12 months, representing 42.5% of all the answers.

There were some differences in the prescription intentions between the 2 professional groups analyzed. For this variable,
psychologists revealed a 48% share of increased likelihood answers, whereas psychiatrists only had 22% of their responses corresponding to these.

Concerning professionals’ attitudes, 136 mental health care professionals (n=114, 84% psychologists; n=22, 16% psychiatrists) declared to have a generally positive (“Positive” or “Very Positive”) attitude toward the possibility of mental health professionals and psychologists being able to prescribe, recommend, or use clinically and technically validated DMHAs, representing 85% of all the answers. A much smaller difference was found in comparison with their intentions to prescribe. Psychologists and psychiatrists responded with 88% and 81%, respectively, generally positive attitude answers.

It is important to note while describing prescription intentions that practically no respondents to the survey have prescribed DMHAs to their patients. Simultaneously, there is no legally established prescription and reimbursement process in Portugal, as described in the Introduction section. This leads us to conclude that what we observe in this sample are the aprioristic perceptions and attitudes toward DMHAs.

Respondents who reported more positive attitudes toward DMHAs (χ²=3.9; P=.048; Cramer V=0.19) and those who worked in a clinic (Fisher exact P=.03) reported a higher intention of prescription. Male respondents also reported a higher likelihood of assuming more positive attitudes toward DMHAs (Fisher exact P=.046). The demographic or work-related characteristics of other health care professionals were not significantly associated with either DMHA attitudes or prescription intentions.

We found a statistically significant association between the digital affinity score and prescription intention (Fisher exact P=.01). This was not the case for the association between the digital affinity and the mental health professionals’ attitudes (Fisher exact P=.67). Our results lead us to believe that Portuguese professionals expect to prescribe DMHAs shortly (the next 12 months); however, they are not currently very optimistic about these tools. This may suggest that they are open to changing their views.

Perceived Barriers to DMHA Prescription

Overall, 11 potential obstacles to DMHA prescription were listed. For most cases, except for lack of support from the manufacturer for technical issues, the answers demonstrated an agreement or total agreement with the characterization of the following as barriers to the prescription of these solutions.

With the highest share of agreement (154/160, 96.3%), the lack of information about digital apps gathered the most support from the respondents, including the highest percentage of “Totally Agree” answers. In addition, important issues such as the initial effort for health professionals (115/160, 71.9%), the need to adjust and adapt clinical records and practices (113/160, 70.6%), and ethical and legal questions (110/160, 68.8%) were very commonly identified as obstacles.

Respondents were found to disagree more frequently with the idea that uncertainty around informed consent (33/160, 20.6%) and data privacy and safety (28/160, 17.5%) would be substantial obstacles to DMHA adoption. At the same time, the most neutral answers (“Don’t Know” or “Neither Agree nor Disagree”) were registered regarding the lack of support from the manufacturer for technical issues (92/160, 57.5%), considerably superior to the next highest value (absence of copayment mechanisms; 56/160, 35%).

Measures to Support Adoption of DMHAs

In total, 10 different measures were presented to increase the adoption of DMHAs. For all cases, respondents expressed a favorable agreement with the utility of their adoption, ranging from 52% to 94% of the answers.

The highest share of agreement was reached concerning the need for more information about available DMHAs (151/160, 94.4%). Closely behind, most professionals considered that the existence of scientific evidence about the validity of the apps (147/160, 91.9%), the recommendation by professional and scientific societies (145/160, 90.6%), and the definition of the prescription process (138/160, 86.3%) along with the sharing and reporting of positive experiences by peers (137/160, 85.6%) were relevant measures to foster adoption.

The highest levels of disagreement were registered for the need to integrate DMHAs in health insurance plans (23/160, 14.4%), followed by changes to the legal framework (10/160, 6.3%) and manufacturer helplines for health care professionals (8/160, 5%). The most relevant share of neutral answers was registered regarding the integration of DMHAs in health insurance plans (54/160, 33.8%), which also reported the lowest level of agreement (83/160, 51.9%).

KOL Analysis

Overview

Of the 25 contacted KOLs, 11 (44%) replied to our invitation to participate in the interviews, 7 (64%) of whom gave positive replies. One KOL did not show up for the scheduled interview, and the remaining 6 were interviewed during the period mentioned in the Methods section. The interviews started with an overview of the study by the coauthors. They proceeded to ask the KOLs what were, in their opinion, the 3 main benefits of, barriers to, and measures to support the adoption of DMHAs in the Portuguese context in descending order. An initial briefing was shared with the invitation to participate.

Perceived Potential Benefits From DMHAs and Attitudes Toward DMHAs

The 3 main benefits identified by the interviewed KOLs and that gathered consensus were the following:

- Improved access and accessibility to health care at the personal and population levels, including geographically more remote areas and preventive services.
- Improved efficiency in providing care, both from the point of view of direct cost (payment per treatment) and allocation of available human resources.
- Proximity to the user (including personalization of care, real-time monitoring, consideration of the user as an active participant, a more relevant number of potential users, and a potential for collecting real-world data for research).
Perceived Barriers to DMHA Prescription

The 3 main perceived barriers to DMHA adoption identified by interviewed KOLs and that gathered consensus were the following:

- Lack of knowledge and literacy of professionals at the digital skills level about existing DMHAs and related evidence-based information.
- Absence of training programs on DMHAs, especially those that positively position them and do not constitute them as a threat to professionals (especially as a risk of being replaced).
- The health system is not designed to consider digital tools (including their use, technical standards for information technology and information security, reimbursement, and stabilization of ethical and data protection concepts).

Measures to Support Adoption of DMHAs

The 3 main measures to support DMHA adoption identified by interviewed KOLs and that gathered consensus were the following:

- Creation of appropriate regulation, especially for clinical practice and reimbursement, and adequate health policies to boost the digital component safely.
- Promotion of literacy on mental health and training in digital apps (especially the younger generations), building awareness, and competency in digital tools in a constructive and collaborative perspective.
- Production of directives by the Ministry of Health and Professional Orders, such as an executive document with a selection of apps that could be useful and easy to implement while demonstrating good screening test characteristics.

Input Gathered From the KOLs

The second part of the interview consisted of the presentation of the results of the survey (as displayed in the previous subsections of the Results section for the web-based survey, ie, Demographics, Perceived Potential Benefits From DMHAs and Attitudes Toward DMHAs, Prescription Intentions of DMHAs, Perceived Barriers to DMHA Prescription, Measures to Support Adoption of DMHAs) to the KOLs and asking them whether they believed the obtained results to be aligned with their perception of most Portuguese psychiatrists and psychologists, justifying why.

Concerning the benefits of DMHAs, KOLs were aligned with the identified benefits for both users and professionals, as well as with the answer distribution of the sample. No KOL dissented from this view. KOLs expressed a perception of bias in the sample, reflecting a higher proportion of promoters of digital tools than the global average of mental health professionals. The answer to the hypothesis “improved ability to make informed choices” was the point that raised the most questions, and it is interesting to understand the rationale behind it. Greater adherence to treatment, access, disease management, and health literacy were in line with expectations. Response data were more balanced when evaluating the benefits for health professionals than for patients, where positive expectations seem to exist a priori.

Concerning the perceived barriers to the adoption of DMHAs, KOLs were generally aligned with identified benefits and the distribution of the answers in the survey, with only 1 KOL stating that they did not think these were representative of their peers’ opinions. KOLs considered that respondents had a favorable perception of digital apps, with a lack of technical support from the manufacturer and the need to adjust work processes surprising them. The first item was a surprise owing to the low degree of disagreement; the second item surprised them because it entailed that the adaptation process would necessarily be painful. In the KOLs’ opinion, professionals need help formulating the problems that concern them the most and whether or not it is a problem. They attributed these issues to a lack of experience with DMHAs. Furthermore, KOLs stated that they expected that a higher percentage would agree with the lack of reimbursement owing to being included in insurers’ commercial packages as a relevant barrier.

Finally, regarding the measures to support the adoption of DMHAs, the KOLs aligned with the adoption support measures identified in the survey and the distribution of responses. The action “integration of applications in health insurance” generated the most comments. KOLs considered that it could reflect ≥1 of the following 3 issues: a priori concerns about data privacy (including data sharing with third parties), low payment fees to professionals, or matters related to stigma.

In addition, KOLs pointed out several peculiarities of the Portuguese context that they believed were important for any stakeholder (government, business, academia, or others) who wishes to develop a likely successful DMHA to address mental health professionals’ needs. These were as follows:

- Health system financing and the incentives it produces must be considered, as health systems with budgeting practices based on production estimates instead of outcomes will experience severe difficulties in monetizing DMHAs and validating them as productive or cost-effective investments.
- Even if they are not inferior to other interventions, DMHAs may allocate resources more effectively and deliver savings by shifting individuals with lower mental health care needs to DMHAs and allowing mental health professionals to focus more of their time and attention on more complex cases.
- Professionals’ resistance to novelty and workflow change, as well as negativity bias and feelings of being replaced by apps, must be addressed to ensure a successful embracement of DMHAs. KOLs considered this to be particularly true for psychiatrists, supported by the number of people who answered that they had the expectation of increased initial effort for health professionals (72% of the survey’s respondents), need to adjust and adapt clinical records (71%), and additional workload (56%).
- DMHAs must be adapted for use in a clinical setting, namely for severe mental issues (such as schizophrenia), where DMHAs are currently unfit to deal with acute severe episodes. Furthermore, DMHAs must be balanced to prevent a user’s perception of pseudautonomy that leads to the
early abandonment of therapeutic interventions. Further research and development are required in these areas.

- Stigma plays a key role. This is not exclusive to one type of actor and ranges from the perceptions that professionals have of users and patients to insurance companies’ pricing policies and offers. Information sharing with third parties other than professionals and users must be selective and scrutinized to prevent distrust in these tools and to avoid discrimination toward people who use them.
- Any DMHA must bear in mind cybersecurity risks and their impact on the user. Mental health issues are usually intimate matters, and that places a higher emphasis on information security.
- Research and development on DMHAs must be ongoing, both from a clinical and a technical standpoint. Evidence generation, treatment, and analysis are expected to be performed on a rolling basis owing to their digital nature.

Discussion

Principal Findings

Perceived Benefits, Barriers, and Measures to Foster Adoption

Regarding the perceived benefits of DMHAs for patients, our findings were generally aligned with those of the study by Dahlhausen et al [55] despite notable differences. Portuguese respondents were less optimistic about improved access to care and more positive about enhanced adherence to treatment. The time-saving potential benefit generated the most positive responses, whereas it was the worst regarding the perceptions of shared questions in the study by Dahlhausen et al [55]. The same is true for patient satisfaction, which was almost evenly split in the study by Dahlhausen et al [55]. Conversely, Portuguese professionals had some of the least positive perceptions concerning treatment success and quality of patient assistance. In contrast, German professionals ranked some of their most positive scores for this variable. Overall, participants demonstrated a very high agreement with the listed potential benefits.

The respondents were less enthusiastic about the perceived benefits of DMHAs for professionals, namely, regarding the expectation of improved treatment success and the possibility of improving the quality of patient care. They did recognize the benefits of better time management owing to efficiency gains and the benefit of having additional treatment options.

We found a wider gap between the practitioners’ attitudes toward DMHAs and their intentions to prescribe them in comparison with German professionals. This may be because of the effects of social desirability bias on provided answers [63-66] and the consequent positive aprioristic expectations. Their role might be expanded in the provided answers given the absence of a regulatory track for prescription and payment and general knowledge about DMHAs.

Concerning the barriers to DMHA adoption, our respondents agreed more with the importance of the lack of information about digital apps and the initial effort for health professionals, as well as the need to adjust and adapt clinical records and practices, alongside ethical and legal questions. They demonstrated general neutrality toward the importance of the lack of support from the manufacturer for technical issues and reimbursement schemes, and KOLs have attributed this neutrality to a lack of knowledge about these tools or concerns with patient data privacy.

By contrast, these mental health professionals agreed that more information about DMHAs, increased scientific evidence about their validity, recommendations by professional and scientific societies, and the definition of a prescription process along with the sharing and reporting of positive experiences by peers were all relevant measures to foster adoption.

To leverage DMHA adoption, both Portuguese and German professionals recognized the importance of the first 2 points and concurred on classifying direct exchange with developers as one of the least important issues. They disagreed, however, on the necessity of integrating apps into health insurers’ commercial packages.

Attitudes and Prescription or Recommendation Intentions

In the German study, health care professionals with higher digital affinity were considerably more positive toward attitudes and prescription intentions; however, the strength of the associations was weak. While DiGA is already at work, DMHAs do not currently have a clear path for partaking in Portugal’s clinical process and care provision.

Although a direct comparison between the professional groups of this study and the ones in the replicated paper is not immediate, the strictest association occurs in practitioners in psychiatric specialties (“child and adolescent psychiatry and psychotherapy” and “psychiatry and psychotherapy”), corresponding to the Portuguese psychiatrists, and the remaining specialties (“psychological psychotherapy” and “psychosomatic medicine and psychotherapy”), corresponding to psychologists in the Portuguese case.

The first group’s differences between attitude and intention toward recommendation or prescription ranged from 38.5% to 45.9%, whereas the second group’s differences ranged from 35.2% to 36.6%. In our study, for psychiatrists, we found a difference of 59% between the reported positive attitudes toward DMHAs and their intention to prescribe them (between 81% and 22%, respectively). For psychologists, the same difference was 40% (between 88% and 48%, respectively).

Although Portuguese psychologists’ answers are more favorable than their German counterparts, the differences between attitudes and intentions have a similar magnitude. However, more Portuguese psychiatrists presented positive attitudes, while slightly fewer reported prescription intentions, thus yielding a larger difference than the one found in German psychiatrists.

KOL Analysis

First, it is important to highlight that in a country with no payment or reimbursement tracks or clinical or technical validation standards specific to digital health apps, these findings are based on the individual KOL’s experience and perception of the national landscape.
Second, the interviewed KOLs mainly considered that the survey’s results, despite the sample size and a possible bias in favor of digital tools, were representative of the study’s intended population. This increased our confidence in the obtained results and, consequently, in the conclusions they can draw toward DMHA promotion.

Third, the fact that KOLs were unanimous regarding several issues—such as access to care, patient centricity, or (need for) mental health professionals’ digital literacy draws attention to the fact that the Portuguese health system needs profound transformation. Alerts have been abundant since Europe’s health systems suffered from the COVID-19 pandemic and its clinical backlog [67], which have come on top of long-lasting struggles such as workforce skills, motivation, and retention [68,69]; speed of digital transformation [70]; or the need to foster innovation [71]. The reference to these topics in unison between knowledgeable people, in a blinded and independent fashion, strengthens these arguments and reinforces the need to act on a broader digital health strategy that encompasses digital health applications.

Relationship With the Wider Portuguese Health System

The reported results from the survey and the KOL analysis revealed a general immaturity in implementing DMHAs (digital medical products, services, and interventions in general) in Portugal.

The survey shows an explicit generalized agreement with the perceptions listed concerning potential benefits, barriers, or measures to foster adoption. These results seem polarized to one of the extremes, as disagreement answers were never >21% for the specific questions.

Moreover, the topics generating the most neutral answers require some practical implementation of these tools—specifically, on the improvement of patient satisfaction, treatment success, and capability to make informed choices; on the lack of support from the manufacturer for technical issues; and on reimbursement of medical prescription, as well as integration of DMHAs in health insurance coverage. We hypothesize that this neutrality confirms a lack of practical experience with these tools. Otherwise, professionals would have more positive or negative perceptions because of their experience and less ambiguity or one-way polarization.

Our theory is compatible with other findings from the survey; although Portuguese professionals are at least as positive in their attitudes toward DMHAs as German professionals, the former group has only a smaller share of those who do intend to prescribe them, thus generating a wider gap between attitudes and prescription intentions for Portuguese mental health professionals than that found for German counterparts in the study by Dahlhausen et al [55]. If their intentions to prescribe are inferior to their German counterparts, then it is expected that they will do so less often, aggravating the know-difference in the practical knowledge of using these tools. They also differ regarding the perceived importance of required workflow adjustments as a limitation to adoption.

This theory is also compatible with the Portuguese health system paradigm concerning mental health as described by the KOLs, who repeatedly reported the professionals’ resistance to change and novelty, as well as the intrinsic fear of being replaced by DMHAs, as barriers to DMHA adoption. This barrier should be taken seriously so as not to create a negative reinforcement loop that further restrains professionals from adopting and applying such tools.

This is all the truer as the Portuguese National Health Plan [72] ranks “access to mental healthcare” as the sixth most important health determinant for the country’s needs but fails to set any of its 37 health objectives to address this issue and its consequences. A specific National Program for Mental Health, responsible for producing a National Plan for Mental Health, existed from 2008 [73] until 2020. The National Coordination for Mental Health Policies replaced it at the beginning of 2022 [74]. Although the former has considerably failed to enact mental health care reform [75] and had no mention of DMHAs, the new Coordination is yet to publish its plan and objectives.

Strengths and Limitations

Portugal faces important challenges despite being a relatively small country (approximately 10.5 million inhabitants), namely, a rapidly growing aging population [76], declining birth rates [77], and an overburdened health system [46]. These challenges are similar to those encountered in many high-income countries. Therefore, Portugal might serve as a test bed to validate digital solutions that ease the workload on health care providers and increase patient autonomy. Our study contributes to this understanding as it is, to the best of the authors’ knowledge, the first work of its kind on the Portuguese landscape concerning DMHAs.

The study elicits preferences and issues that are clearly important to understand the demand side, as it is visible, for example, through the high degree of agreement in survey answers and KOL responses. Furthermore, the mixed methods methodology allows us to combine the perspective of those closest to potential users with a helicopter view with in-depth knowledge, making the derived conclusions more robust.

All biases inherent to sampling and KOL selection are potential limitations of this study, with the social desirability of the provided answers and the role of expectation in the survey’s answer process (ie, the belief respondents might have that they are expected to answer more favorably about technology than they would otherwise do), as well as self-selection, being among the most relevant. Their impact on provided answers is mitigated by the fact that in a postpandemic reality, telehealth (albeit limited to teleconsultation volume) [53] has proven its benefits to a large extent and certainly more than at the time Dahlhausen et al [55] conducted their study.

The sample number may constitute a further limitation, as 160 answers only partially characterize a population of 1528 psychiatrists and an undetermined number of psychologists. However, the level of agreement between the elements gathered in the survey and the KOLs’ assessment, both before and after seeing the survey results, leads us to believe that the underlying uncertainty in the provided answers might not be as considerable.
as expected. Furthermore, the proportion between psychologists and psychiatrists in our study and the study by Dahlhausen et al [55] is approximately the same (6 to 1, respectively). This is a relevant sample characteristic that favors the comparison between the results of both studies.

Finally, the comparability between studies is limited by their different scopes and stages of regulatory development: the study by Dahlhausen et al [55] was produced when DiGA was starting to roll out and for all digital health applications (ie, not exclusive to mental health); as of the conclusion of this study, Portugal does not have a payment or reimbursement system in place for digital health apps in general, or DMHAs specifically. Even then, the similarity of methods, presentation of results, and the produced discussion allow, in our view, for a proper comparison.

Future Work

From a general perspective of digital health apps, it would be useful to fully replicate this study to understand which points are common to all apps and which are solely applicable to DMHAs. Deloitte’s report for Health Cluster Portugal [78] is the only known work on the Portuguese domestic market. However, its high-level nature reveals how immature the market is in terms of digital health apps. More research is needed to understand their market dynamics, namely, when it comes to the expectations of supply and demand sides. To further improve the interpretation of results and have a clearer sense of the actual differences between Portuguese mental health professionals and their German counterparts, it would be helpful to implement this or similar surveys on the overall Portuguese population.

By evaluating the differences between mental health professionals and the people they serve, one could distinguish between context and specific differences attributable to training or skillset.

Moreover, as this is a portrait of the landscape as far as DMHA are concerned and one of the main difficulties that we felt was survey engagement (despite extensive dissemination efforts), it would be essential to perform this extended work with further engagement from professional orders, professional societies, and market-based stakeholders (eg, developers of apps such as 29kFIN [79]). Outreach to international stakeholders such as BfArM, mobile health Belgium, and market operators with products that have already secured regulatory approval by them would be beneficial.

Given that this necessarily entails regulation, it would be interesting to expand further on the regulatory science angle and the opportunity for either regulatory tracks at the EU level (following calls for European Health Union and increasing competencies of the EU in health care [80]) or regulatory replication to ensure a fair, competitive, and innovative digital health market in the EU. This is in the interests of regulators, companies, and citizens.

Finally, it is widely agreed that digital interventions have appreciable potential to deliver more and better health care. However, if these—or any other health interventions—are to succeed, they need rigorous planning by diagnosing the status, defining targets and priorities, establishing objectives and desirable results, and outlining the best evidence-based strategies and plans to achieve them [81]. Monitoring and evaluating the attained results and deriving learned lessons are also necessary. In the authors’ view, this road map for DMHAs would potentially be the most relevant future work that could be done.

Conclusions

Portuguese mental health professionals’ perceptions of digital health apps present clear aprioristic expectations regarding the benefits of DMHAs for users, especially concerning improved therapeutic adherence and health literacy. Although professionals generally recognize the benefits for patients, they are less optimistic about the expected advantages for themselves and their peers. Although the usefulness of DMHAs for professionals needs to be clarified, benefits such as efficiency gains and having an additional treatment option are among the most valued benefits from the onset.

Chief among the main perceived barriers are the need for more information about digital health apps, preconceptions of initial use efforts for health professionals, and the need to adjust and adapt clinical records. The main enablers of DMHA use identified include more information about these apps, both regarding how they work and scientific evidence about the validity of such apps, as well as recommendations by professional and scientific societies. Governmental or regulatory guidelines are strongly recommended.

Portuguese mental health professionals, compared with German mental health professionals, were similar in most of the reported answers. Some notable differences were fewer positive perceptions concerning treatment success and quality of patient assistance, a wider gap between attitudes and prescription intentions for Portuguese mental health professionals, and the need for considerable workflow adjustments as limitations for adoption.

Concerning how digitally literate mental health professionals perceive themselves and their patients to be, the scores of digital literacy–related issues in terms of barriers and measures to support adoption in the survey, along with the conclusion by the KOLs that this is one of the main issues faced by mental health professionals in Portugal, lead us to posit that mental health professionals perceive themselves to have high degrees of digital illiteracy. They also perceived a strong need for patients to be educated should DMHAs or other digital tools be implemented to deliver mental health care.

Mental health professionals believe that their role in digitalizing health care provision consists mainly of promoting literacy among peers, namely, to and by younger age groups, thus forming communities able to capacitate a growing number of professionals. Their participation in professional and scientific societies is another avenue for further engagement. Finally, mental health professionals believe that the Portuguese government sector should play a crucial role in shaping the health system and enabling the proper organizational and financial means and incentives to catalyze transformation.
Acknowledgments

This work was conducted under the scope of and funded by the Health Data Science PhD Program of the Faculty of Medicine of the University of Porto, Portugal [82].

The authors would like to thank Knok healthcare, the Order of Portuguese Psychologists, the Portuguese Society of Psychiatry and Mental Health, and all those who contributed to disseminating the survey in appropriate media and forums.

The authors would like to thank those whose inputs have made this paper richer in content and form.

Data Availability

The data sets generated or analyzed during this study are not publicly available but are available from the corresponding author upon reasonable request.

Conflicts of Interest

This study was conducted as part of the doctoral research project at the Health Data Science PhD Program, Faculty of Medicine of the University of Porto, Portugal. Two authors participated in this project while employed by the Faculty of Medicine of the University of Porto, as DN-L and RC-C were working on a project cofunded by the European Union concerning the digitalization of clinical trials in the north of Portugal (CR—Digital: Digitize clinical research in the North of Portugal NORTE-01-0145-FEDER-083448). The European-funded project or the Faculty of Medicine of the University of Porto were at no point involved in the research, aside from the required administrative approvals. There was also no funding, pay, or other commercial interest provided by the Faculty of Medicine of the University of Porto, aside from the costs related to the publication of this paper. The help received from Knok healthcare regarding the dissemination of the survey questionnaire entailed no reception or use of financial benefits or otherwise. All work concerning this study was conducted during the authors’ personal time.

Multimedia Appendix 1
Translated adapted survey questionnaire (Portuguese to English).
[DOCX File, 631 KB - humanfactors_v10i1e45949_app1.docx]

Multimedia Appendix 2
Translated original survey questionnaire (English to Portuguese).
[DOCX File, 108 KB - humanfactors_v10i1e45949_app2.docx]

Multimedia Appendix 3
Summary of modifications between survey questionnaires.
[DOCX File, 26 KB - humanfactors_v10i1e45949_app3.docx]

References


23. mHealthBelgium. URL: https://mhealthbelgium.be [accessed 2022-12-01]


Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Survey
DIGA: Digitale Gesundheitsanwendungen
DMHA: digital mental health app
EU: European Union
KOL: key opinion leader

©Diogo Nogueira-Leite, José Miguel Diniz, Ricardo Cruz-Correia. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 02.06.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Abstract

Background: Over the last decade, there has been an increase in the number of health care consumers (ie, patients, citizens, and laypeople) with access to their laboratory results through portals. However, many portals are not designed with the consumer in mind, which can limit communication effectiveness and consumer empowerment.

Objective: We aimed to study design facilitators and barriers affecting consumer use of a laboratory results portal. We sought to identify modifiable design attributes to inform future interface specifications and improve patient safety.

Methods: A web-based questionnaire with open- and closed-ended items was distributed to consumers in British Columbia, Canada. Open-ended items with affinity diagramming and closed-ended questions with descriptive statistics were analyzed.

Results: Participants (N=30) preferred reviewing their laboratory results through portals rather than waiting to see their provider. However, respondents were critical of the interface design (ie, interface usability, information completeness, and display clarity). Scores suggest there are display issues impacting communication that require urgent attention.

Conclusions: There are modifiable usability, content, and display issues associated with laboratory results portals that, if addressed, could arguably improve communication effectiveness, patient empowerment, and health care safety.

Introduction

Ambulatory care practitioners (ie, primary care providers and medical specialists) frequently order laboratory tests for patients as part of a diagnostic evaluation or to monitor the progression of chronic illness [1]. In the past, practitioners received paper reports with test results, whereas now, it is more common to review results in the electronic health record (EHR). Commercial EHR developers assume the target users have sufficient domain expertise to access and use this information with little additional context or instruction.

However, health care consumers (ie, patients, citizens, and laypeople) are increasingly accessing their own laboratory results (eg, COVID-19, Papanicolaou smear, and blood work results) through independent laboratory portals or patient-facing portals tethered to an organizational EHR. Research shows that people want access to their laboratory results [2-4] to track their health status and guide decision-making [4,5]. Those with...
chronic illnesses can use this information to monitor [6] and more effectively self-manage their own medical conditions [4,7]. Having direct access to laboratory results through portals often means getting results sooner [3,4] and without the inconvenience of scheduling a follow-up appointment or traveling to see a health care practitioner [6]. Furthermore, consumers with results in hand are empowered to engage more effectively in a discussion with clinicians during appointments [2,5] and ensure results are not overlooked [4,6].

It is imprudent, however, to equate access with value; just because consumers can see information does not mean they can understand or use it. Several studies of consumer portals have found that while Canadians appreciate being able to access information on the internet, they struggle to understand and use their results [3,8]. The introduction of new technologies can alter traditional workflows. Circumventing in-person appointments—and the explanations or education practitioners provide during these encounters—may limit communication effectiveness or produce unintended consequences. Laboratory results and technical reports can be complicated and difficult to interpret without medical expertise and additional context.

Researchers have identified several shortcomings in the display of laboratory results portals that reduce their usefulness as patient communication, education, or self-management tools. For example, Leckart [9] noted that laboratory results are typically long text-based reports with many unfamiliar acronyms. The reports also separate patient values from associated reference ranges. By contrast, using graphs to depict values with reference ranges improves consumers’ ability to interpret results [10]. While the need to redesign laboratory reports was well documented over a decade ago [9], little progress was made to include emphasis cues, contextual information, or hypertext links to related resources. Consumers want their test results combined with actionable information [11,12]. Unfortunately, laboratory results portals rarely include context-sensitive interpretation [11] or recommendations to improve values [9]. These challenges are compounded for Canadians with low health literacy [13]. In light of these issues, it is unsurprising that nearly half (46%) of consumers turn to the internet to find answers to their questions about laboratory results [11].

Further complicating matters, consumers may need to use multiple different portals to review all their information. Health care provider organizations within a community of practice may use a tapestry of different EHR vendors, products, and features. Consequently, not only are the data fragmented between different systems, but users may have different user experiences (eg, button locations and information displays) depending on where the laboratory tests were ordered or processed. In Canada, some consumers have had access to some of their laboratory results for over a decade. For example, in British Columbia, independent laboratory results have been on the internet since 2010 [3]. Canadians may now also access laboratory results performed during hospitalizations using patient portals tethered to an EHR. However, the information remains siloed; laboratory results are only accessible through the portal linked to where the laboratory tests were done (ie, ambulatory independent laboratory vs hospital). Therefore, the information is fragmented for users.

The purpose of this study was to identify interface usability issues and associated modifiable attributes of laboratory results portals (including more comprehensive portals tethered to EHRs). Our goal was to uncover design strategies that might inform future portal specifications and improve the communication of laboratory results. We provided a questionnaire to a sample of Canadians asking about their use of laboratory results portals, their perceptions of existing portals, and their perspectives on the design of information displays. Our inquiry was focused on general features related to laboratory results portal systems rather than a specific vendor, product, or health provider organization.

Methods

Study Design

We recruited people by posting an invitation on a web-based platform for health research volunteers in British Columbia, Canada. Participants accessed the questionnaire using a hypertext link; administration was unmoderated. To be eligible, participants (1) needed to have experience using at least one laboratory results portal and (2) be at least 19 years old. Health care professionals or trainees were excluded. Participants were offered CAD $5 (approximately US $3.70) as an honorarium to participate. The questionnaire was available from November 2020 to February 2021 (Multimedia Appendix 1).

In addition to gathering demographic information (eg, age, country of birth, and primary language spoken at home), we asked consumers about their experiences using laboratory results portals (eg, how long they had been using laboratory results portals) and their perceptions of the user experience (eg, usability, understandability, and information needs). Closed-ended Likert-type questions were used (1 star to 5 stars) to measure perceptions of user-friendliness (ie, usability), the available information (ie, content), and display formatting. We also included one question asking, “Would you suggest someone else use a lab results portal?” or a modified Net Promoter Score (NPS) [14] using a 5-star scale. Open-ended (ie, free-form response) questions were included to encourage participants to provide additional context or explanation. Participants were asked to provide general comments about laboratory results portals and specific suggestions for improving (1) the use of laboratory results portals, (2) the information (ie, numbers and words) contained in laboratory results portals, and (3) the laboratory results portal displays (eg, color and format).

Ethics Approval

The University of Victoria’s Human Research Ethics Board approved this study (20-0712).

Quantitative and Qualitative Analysis

Descriptive statistics were used to examine responses to the closed-ended questions. The modified NPS was calculated by first categorizing responses and calculating percentages for each category (ie, 1-3=detractors, 4=passive, and 5=promoters). We
then subtracted the percentage of detractors from the percentage of promoters [14].

To analyze the qualitative data, we used affinity diagramming (ie, affinity mapping) to identify commonalities between responses [15,16]. Affinity diagramming is a common qualitative research method used to organize findings (eg, comments and observations) into groups that share semantic meaning or concepts [16]. The researchers (HM and LM) met over the web using Zoom videoconferencing software (Zoom Video Communications) to screen share and Microsoft PowerPoint to visualize and categorize each participant’s response. The responses to each open-ended question were analyzed separately. In cases where a participant’s response contained more than one concept, we separated the response into as many independent concepts as necessary. Each concept was also color-coded to indicate whether the content was positive, negative, or a suggestion for improvement. The groups of comments that emerged, reflecting the thematic similarities, were named. Some of these categories were hierarchical with subcategories. Finally, the content in each category was synthesized into a summary description.

After affinity diagramming, we compared our inductively coded categories to themes in the literature [17]. We replicated the coding from the affinity diagramming using MaxQDA (Verbi) qualitative analysis software to count the frequency with which each category was mentioned by participants.

Results

Descriptive Statistics

In total, 30 people completed the questionnaire (ie, N=30). Most participants were between 45 years or older (n=17, 57%), women (n=26, 87%) and born in Canada (n=23, 77%); spoke English at home (n=29, 97%); and had at minimum some post graduate training (eg, certificate, Bachelor’s degree) (n=16, 53%) (Table 1). Nearly three-quarters (n=22, 73%) of the participants had at least 1 chronic condition. The most common conditions reported included cardiovascular disease (n=6, 20%), mental illness (n=7, 23%), and musculoskeletal disorders (n=7, 23%). Most participants (n=25, 83%) took one or more prescription medications in the past 2 days.

There was variability in the amount of laboratory tests respondents had done and their use patterns of laboratory results portal use (see Table 2).
Table 1. Demographic characteristics of the sample.

<table>
<thead>
<tr>
<th>Demographic characteristic</th>
<th>Value, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
</tr>
<tr>
<td>19-24</td>
<td>1 (3)</td>
</tr>
<tr>
<td>25-34</td>
<td>8 (27)</td>
</tr>
<tr>
<td>35-44</td>
<td>4 (13)</td>
</tr>
<tr>
<td>45-54</td>
<td>7 (23)</td>
</tr>
<tr>
<td>55-64</td>
<td>6 (20)</td>
</tr>
<tr>
<td>65-74</td>
<td>4 (13)</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>26 (87)</td>
</tr>
<tr>
<td>Men</td>
<td>3 (10)</td>
</tr>
<tr>
<td>Prefer not to disclose</td>
<td>1 (3)</td>
</tr>
<tr>
<td><strong>Country of birth</strong></td>
<td></td>
</tr>
<tr>
<td>Canada</td>
<td>23 (77)</td>
</tr>
<tr>
<td>Other</td>
<td>7 (23)</td>
</tr>
<tr>
<td><strong>Primary language spoken at home</strong></td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>29 (97)</td>
</tr>
<tr>
<td>Other</td>
<td>1 (3)</td>
</tr>
<tr>
<td><strong>Highest level of education</strong></td>
<td></td>
</tr>
<tr>
<td>Secondary (high) school diploma or equivalency certificate</td>
<td>2 (7)</td>
</tr>
<tr>
<td>Trades certificate or diploma other than Certificate of Apprenticeship or Certificate of Qualification</td>
<td>3 (10)</td>
</tr>
<tr>
<td>College, College of General and Professional Teaching, or other nonuniversity certificate or diploma</td>
<td>4 (13)</td>
</tr>
<tr>
<td>University certificate or diploma below bachelor level</td>
<td>5 (17)</td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td>8 (27)</td>
</tr>
<tr>
<td>University certificate or diploma above bachelor level</td>
<td>2 (7)</td>
</tr>
<tr>
<td>Master’s degree</td>
<td>5 (17)</td>
</tr>
<tr>
<td>Doctoral degree</td>
<td>1 (3)</td>
</tr>
<tr>
<td><strong>Chronic illnesses</strong></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>22 (73)</td>
</tr>
<tr>
<td>No</td>
<td>8 (27)</td>
</tr>
<tr>
<td><strong>Type of chronic illnesses</strong></td>
<td></td>
</tr>
<tr>
<td>Musculoskeletal disorder</td>
<td>9 (30)</td>
</tr>
<tr>
<td>Neurological condition</td>
<td>8 (27)</td>
</tr>
<tr>
<td>Mental illness</td>
<td>8 (27)</td>
</tr>
<tr>
<td>Cardiovascular disease</td>
<td>6 (20)</td>
</tr>
<tr>
<td>Chronic respiratory disease</td>
<td>5 (17)</td>
</tr>
<tr>
<td>Cancer</td>
<td>3 (10)</td>
</tr>
<tr>
<td>Chronic pain</td>
<td>3 (10)</td>
</tr>
<tr>
<td>Diabetes</td>
<td>2 (7)</td>
</tr>
<tr>
<td>Other</td>
<td>8 (27)</td>
</tr>
<tr>
<td><strong>Number of prescription medications taken in the past 2 days</strong></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>5 (17)</td>
</tr>
<tr>
<td>1</td>
<td>5 (17)</td>
</tr>
</tbody>
</table>
Table 2. Descriptive statistics for participant use patterns.

<table>
<thead>
<tr>
<th>Descriptive characteristic</th>
<th>Value, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Last laboratory test done</strong></td>
<td></td>
</tr>
<tr>
<td>In the past month</td>
<td>17 (57)</td>
</tr>
<tr>
<td>In the past 6 months</td>
<td>7 (23)</td>
</tr>
<tr>
<td>In the past year</td>
<td>5 (17)</td>
</tr>
<tr>
<td>In the past 5 years</td>
<td>1 (3)</td>
</tr>
<tr>
<td><strong>Frequency of getting laboratory tests done</strong></td>
<td></td>
</tr>
<tr>
<td>A few times a month</td>
<td>2 (7)</td>
</tr>
<tr>
<td>A few times a year</td>
<td>21 (70)</td>
</tr>
<tr>
<td>Once a year</td>
<td>5 (17)</td>
</tr>
<tr>
<td>Less than once a year</td>
<td>2 (7)</td>
</tr>
<tr>
<td><strong>Laboratory results portals used</strong></td>
<td></td>
</tr>
<tr>
<td>Myehealth.ca (since renamed MyCareCompass)</td>
<td>28 (93)</td>
</tr>
<tr>
<td>Other</td>
<td>8 (27)</td>
</tr>
<tr>
<td><strong>Started using laboratory results portals</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;1 year</td>
<td>1 (3)</td>
</tr>
<tr>
<td>2-3 years</td>
<td>13 (43)</td>
</tr>
<tr>
<td>4-5 years</td>
<td>6 (20)</td>
</tr>
<tr>
<td>5+ years</td>
<td>10 (33)</td>
</tr>
<tr>
<td><strong>Frequency of using laboratory results portals</strong></td>
<td></td>
</tr>
<tr>
<td>A few times a month</td>
<td>5 (17)</td>
</tr>
<tr>
<td>A few times a year</td>
<td>23 (77)</td>
</tr>
<tr>
<td>Once a year</td>
<td>1 (3)</td>
</tr>
<tr>
<td>Less than once a year</td>
<td>1 (3)</td>
</tr>
</tbody>
</table>

Most respondents first began using laboratory results portals (n=20, 67%) within the last 5 years and reviewed them several times a year (n=23, 77%). Nearly all respondents (n=29, 97%) reported using an independent laboratory portal available in the province (ie, MyeHealth.ca, which was recently renamed MyCareCompass), with 4 (13%) participants also using portals tethered to hospital EHRs. One (3%) participant reported only using a tethered hospital EHR laboratory results portal but not the independent portal.

**Overall Ratings of Laboratory Results Portals**

Generally, participants rated laboratory results portals favorably but indicated opportunities for design improvements (Table 3). Most participants were very likely to recommend laboratory results portals to others (modified NPS=50; 19/30, 63.3% promoters – 4/30, 13.3% detractors). Participants scored usability the highest, followed by information, and then display. We review each dimension in the next section.
Table 3. Overall ratings of laboratory results portals.

<table>
<thead>
<tr>
<th>Question topic</th>
<th>Question</th>
<th>Value, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modified Net Promoter Score</td>
<td>Would you suggest someone else use laboratory results portals?</td>
<td>4.5 (0.73)</td>
</tr>
<tr>
<td>Usability</td>
<td>Overall, how user-friendly are your laboratory results portals?</td>
<td>3.7 (1.09)</td>
</tr>
<tr>
<td>Information</td>
<td>Overall, how would you rate the information (numbers and words) from laboratory results portals?</td>
<td>3.5 (0.73)</td>
</tr>
<tr>
<td>Display</td>
<td>Overall, how would you rate the display (layout, font size, color, etc) of laboratory results portals?</td>
<td>3.3 (0.60)</td>
</tr>
</tbody>
</table>

Usability and Features of Laboratory Results Portals

Overview
Consistent with the overall usability rating (mean 3.7, SD 1.07; see Table 3), most participants indicated that creating an account (n=23, 77%), logging in (n=29, 97%), and finding information (n=28, 93%) was easy or very easy (Figure 1). Only 1 (3%) participant had difficulty logging into the portal, and 3 (10%) had trouble locating information. Participants had the most difficulty creating an account—4 (13%) said this was hard or very hard.

![Figure 1. Perceived usability of laboratory results portal tasks.](image1)

Features of Laboratory Results Portals
We examined which features participants used (Figure 2). Most had booked an appointment on the internet (n=25, 83%) and used the platform to find a laboratory location (n=22, 73%). Fewer participants (n=18, 60%) had used the analytics page.

![Figure 2. Use of laboratory results portal features.](image2)

A patient can choose to share their report with someone else (eg, a family member or caregiver) by clicking a button and inputting an email address. The patient can also customize user viewing privileges to their health data. More participants reported seeing someone else’s laboratory results (n=10, 33%) than sharing their results (n=4, 13%). None reported changing their results to another language.

Respondents were supportive of the addition of a notification feature. Specifically, all but one respondent (n=29, 97%) wanted a notification to let them know when their results were available in the portal.
Information in Laboratory Results Portals

Despite rating overall portal information positively (mean 3.5, SD 0.73; Table 1), participants were more critical of specific aspects (Figure 3). In total, 21 (70%) respondents found the information easy or very easy to understand, whereas half (n=15, 50%) found it hard to make decisions based on their results.

Figure 3. Perceived understanding and ability to use information from laboratory results portals.

Laboratory Results Portal Displays

Display scores were the lowest overall (mean 3.3, SD 0.60; Table 1). We asked participants about 4 display attributes (Figure 4). In total, 25 (83%) respondents liked or really liked the colors of the laboratory results portal displays; 23 (77%) liked or really liked the layout, and 21 (70%) liked or really liked the font size. Spacing seemed most problematic; only 12 (60%) scored spacing positively.

Figure 4. Ratings of visual aspects of laboratory results portal displays.

Qualitative Themes From Open-Ended Responses

Overview

We identified four major themes with affinity diagramming: (1) overall access, (2) usability and features, (3) information, and (4) displays. To be included as a theme, we set an a priori reporting threshold of 25% (ie, at least 8, 27%, participants had to articulate the theme for us to include and describe it). If a theme could belong to more than 1 category (eg, participants reported issues with the usability and displays of trend feature), we only report it once for brevity.

Overall Access

Many participants (n=11, 37%) liked having access to their laboratory results. Some felt more independent without having to rely upon their health care provider as an “information gatekeeper.” For example, participant 4 wrote, “I love having the option to look it up and not having to wait for my doctor to tell me the results.” Participant 27 wrote, “[It is] helpful to have access to the results, as often doctors don’t let you know what the results are.” Other participants believed this access helped them prepare for appointments, engage in clinical discussions, or manage their conditions. For example, participant 23 wrote, “[this is] essential info to doing my part to manage my health challenges.”

Usability and Features

We explored participants’ insights about several dimensions of usability, including effectiveness, efficiency, and satisfaction in meeting their goals. Some participants believed laboratory results portals were quick, easy, and straightforward to use, whereas others said the portals were hard to navigate. Many also described difficulty finding specific clinical information.

Some participants mentioned wanting a mobile app rather than using a web browser. For example, participant 30 said, “an app would be amazing.” Several (n=8, 27%) also described usability problems when using laboratory results portals on mobile devices. Some respondents expressed difficulty navigating these...
portals or understanding the interface user flow on mobile devices. Participant 21 explained, “the site could have a better flow for mobile. Still easy to navigate, but not pleasing to the eye on the mobile version.”

Respondents also indicated that the addition of a notification feature to let them know when their results were available in the portals would be beneficial to remind them to check, prevent continuous checking, and save time. For example, participant 10 wrote “after a test, I check repeatedly to see if results are ready. It would be convenient if I just received an email when they [results] are ready.”

Two-thirds (n=20, 67%) of our respondents commented on the interface’s ability to display laboratory trends. Many liked that portals allowed them to track their values over time or easily recognize when values were improving or deteriorating. However, others (n=13, 43%) described usability issues. For example, participant 3 said, “I find trying to get the trends over time doesn’t seem to be user-friendly.” Many respondents either did not have the capability to view trends, had difficulty viewing them, or were unaware the feature existed.

Information Needs
Respondents (n=17, 57%) wanted more descriptions about the tests, reference ranges, information about clinical relevance (e.g., creatinine is a measure of kidney function), the meaning of abnormal results, and links to supplementary information. For example, participant 24 wanted “some kind of explanation for people to understand what was being tested and why,” and participant 12 wrote, “[It] would be great to add links to information on what tests are used for and what abnormal results indicate.”

Many participants (n=12, 40%) may have struggled with medical jargon and cited a lack of plain language explanations. Additionally, participants (n=8, 27%) wanted to know what acronyms stood for by providing definitions or including a glossary within the laboratory results portal. For example, participant 17 said, “spell out any abbreviations of test info,” and participant 1 wrote, “I often know what type of tests are ordered, but don’t what the items associated with test means – e.g., under haematology, what is MCV? MCH? MCHC?”

Displays
Participants complained about attributes of the data displays, including how abnormal values are rendered, use of color, and font size. Many participants (n=17, 57%) wanted out-of-range values to be easier to recognize. For example, respondents suggested emphasizing abnormal values with bold font or highlighting. Participant 10 wrote, “[we need] bold or coloured for abnormal results.” In all, 13 (43%) respondents said color could be improved. They suggested using alternate row shading to make results easier to read and including a color scale or coding scheme for out-of-range test results (e.g., red is abnormal and green is within the normal range). When asked about opportunities for improvement, participant 17 wrote, “colour coding: normal, high, low; shade alternating lines to make it easier to read.” In total, 9 (30%) participants wanted larger font or the option to increase the font size.

Concerning communication effectiveness, participants indicated they needed additional information to understand their results and contextualize them to their own health status. In the free-text comments, respondents asked for descriptive text, decoded acronyms, the interpretation of results, qualifying information for abnormal values, and hypertext links to additional resources. This is hardly surprising given the importance of clear communication to overcome health literacy barriers, reduce errors, and improve clinical outcomes [21]. Experts have long advocated universal precautions for health communication when interacting with patients in person or on the internet [22,23]. Universal communication precautions are standard methods for discussing technical information to avoid miscommunication and misunderstanding. This is doubly important when communicating with consumers asynchronously. Therefore, portal designers should take the same steps clinicians are expected to take when engaging patients: using plain “everyday” language, including explanations, checking for understanding, and avoiding information overload through progressive disclosure. During synchronous encounters, experts encourage clinicians to use a “teach back” method (i.e., asking the patient to explain information in their own words) to confirm understanding [24]. This may pose a novel challenge for web developers. Nevertheless, streaming videos, interactive apps, and artificial intelligence chatbots may offer innovative ways to bridge this gap.

Participant responses to our questions about the data display were more critical. It seemed that color, font, spacing, and layout could all be improved to facilitate more efficient information retrieval and more effective understanding. Again, we anticipated this result given the challenges clinicians face when searching the EHR for key laboratory results [25,26]. In usability studies of contemporary EHR interfaces, clinicians searching for diagnostic information have reported difficulty with navigation, data fragmentation, scale interpretation, search functions, and even readability [26]. Moreover, most laboratory reports intended for consumers look similar to those intended for clinicians. Presenting laboratory results to consumers in formats designed for health care professionals is neither helpful...

Discussion
Principal Findings
Overall, most respondents in our study valued having access to web-based laboratory results and were likely to recommend laboratory results portals to other consumers [14]. However, their ratings and estimates of portal attributes, including portal usability, informational content, and data displays, suggest there are usability flaws limiting the quality of use and communication effectiveness. Concerning the quality of use, advanced features, such as data trending and analytics, were absent, difficult to find, or rarely used. This raises the question of whether web-based tools are providing deeper insight into chronic illness management or closing knowledge gaps that may occur if patients bypass discussions with their health care team. This seems to be a lost opportunity since graphic user interfaces offer dynamic tools for data visualization, manipulation, and understanding [18-20].

Concerning communication effectiveness, participants indicated they needed additional information to understand their results and contextualize them to their own health status. In the free-text comments, respondents asked for descriptive text, decoded acronyms, the interpretation of results, qualifying information for abnormal values, and hypertext links to additional resources. This is hardly surprising given the importance of clear communication to overcome health literacy barriers, reduce errors, and improve clinical outcomes [21]. Experts have long advocated universal precautions for health communication when interacting with patients in person or on the internet [22,23]. Universal communication precautions are standard methods for discussing technical information to avoid miscommunication and misunderstanding. This is doubly important when communicating with consumers asynchronously. Therefore, portal designers should take the same steps clinicians are expected to take when engaging patients: using plain “everyday” language, including explanations, checking for understanding, and avoiding information overload through progressive disclosure. During synchronous encounters, experts encourage clinicians to use a “teach back” method (i.e., asking the patient to explain information in their own words) to confirm understanding [24]. This may pose a novel challenge for web developers. Nevertheless, streaming videos, interactive apps, and artificial intelligence chatbots may offer innovative ways to bridge this gap.

Participant responses to our questions about the data display were more critical. It seemed that color, font, spacing, and layout could all be improved to facilitate more efficient information retrieval and more effective understanding. Again, we anticipated this result given the challenges clinicians face when searching the EHR for key laboratory results [25,26]. In usability studies of contemporary EHR interfaces, clinicians searching for diagnostic information have reported difficulty with navigation, data fragmentation, scale interpretation, search functions, and even readability [26]. Moreover, most laboratory reports intended for consumers look similar to those intended for clinicians. Presenting laboratory results to consumers in formats designed for health care professionals is neither helpful...

Discussion
Principal Findings
Overall, most respondents in our study valued having access to web-based laboratory results and were likely to recommend laboratory results portals to other consumers [14]. However, their ratings and estimates of portal attributes, including portal usability, informational content, and data displays, suggest there are usability flaws limiting the quality of use and communication effectiveness. Concerning the quality of use, advanced features, such as data trending and analytics, were absent, difficult to find, or rarely used. This raises the question of whether web-based tools are providing deeper insight into chronic illness management or closing knowledge gaps that may occur if patients bypass discussions with their health care team. This seems to be a lost opportunity since graphic user interfaces offer dynamic tools for data visualization, manipulation, and understanding [18-20].

Concerning communication effectiveness, participants indicated they needed additional information to understand their results and contextualize them to their own health status. In the free-text comments, respondents asked for descriptive text, decoded acronyms, the interpretation of results, qualifying information for abnormal values, and hypertext links to additional resources. This is hardly surprising given the importance of clear communication to overcome health literacy barriers, reduce errors, and improve clinical outcomes [21]. Experts have long advocated universal precautions for health communication when interacting with patients in person or on the internet [22,23]. Universal communication precautions are standard methods for discussing technical information to avoid miscommunication and misunderstanding. This is doubly important when communicating with consumers asynchronously. Therefore, portal designers should take the same steps clinicians are expected to take when engaging patients: using plain “everyday” language, including explanations, checking for understanding, and avoiding information overload through progressive disclosure. During synchronous encounters, experts encourage clinicians to use a “teach back” method (i.e., asking the patient to explain information in their own words) to confirm understanding [24]. This may pose a novel challenge for web developers. Nevertheless, streaming videos, interactive apps, and artificial intelligence chatbots may offer innovative ways to bridge this gap.

Participant responses to our questions about the data display were more critical. It seemed that color, font, spacing, and layout could all be improved to facilitate more efficient information retrieval and more effective understanding. Again, we anticipated this result given the challenges clinicians face when searching the EHR for key laboratory results [25,26]. In usability studies of contemporary EHR interfaces, clinicians searching for diagnostic information have reported difficulty with navigation, data fragmentation, scale interpretation, search functions, and even readability [26]. Moreover, most laboratory reports intended for consumers look similar to those intended for clinicians. Presenting laboratory results to consumers in formats designed for health care professionals is neither helpful...
nor safe. It is critical to ensure outputs are clear and safe to use. Otherwise, consumers may overlook important findings [27] or turn to the internet—and other less scrupulous sources—for help interpreting laboratory results [11,28-31].

Implications of Findings

We believe there is a need for more user experience research with health care consumers to inform the evidence-based design of future patient portals. Developers should test display options to identify what fonts, configurations, colors, and other attributes improve user efficiency, promote action, or reduce errors. This line of inquiry can provide insights into unmet user requirements, new features, and breakthrough innovations. It would also be useful to compare different portals using A/B testing to determine which features or design decisions perform better. For example, do private laboratory portals differ substantially from portals tethered to hospital or clinic EHRs?

We believe that the qualitative responses to our questionnaire offer a base set of requirements for future software development. These quotes can help us to better define problems, understand user needs, and challenge our assumptions. Based on the responses, here is a list of common requirements we believe are important to success: (1) people want timely (ie, quick and on-demand) access to their results; (2) people need access to their data without relying on their providers; (3) people want portals that can be accessed on a browser or personal device; (4) it is important for people to monitor or see trends in their values over time; (5) people want to use results to self-manage their own health conditions; (6) people need assistance interpreting results in the context of their own health; (7) descriptive information should be context sensitive and in plain language; (8) people want clear, easy-to-read displays that highlight abnormal values; (9) abnormal values should include actionable recommendations; and (10) portals should include hypertext links to additional resources or downloads.

These results indicate that in markets where there are alternative options for getting laboratory testing done, it may be prudent for laboratory companies to invest more to attract more customers (consumers). That is, many consumers appreciate laboratory results portals, but the overall user experience could be improved, which could create a competitive advantage. However, if equipping consumers with actionable information leads to them using it to better their health, they may need fewer laboratory tests. Therefore, businesses may actually be deterred from deploying well-designed information this way. However, for countries with public health care, it would be beneficial as it could be used for health promotion, illness prevention, and improved self-management.

Limitations

There were several limitations to this study. First, we recruited a small sample of experienced portal users. Participants were predominantly women; well educated; English speaking; and living in British Columbia, Canada. Therefore, the perspectives presented here may not be representative of other populations.

Second, participants could have used several different laboratory results portals. We did not attempt to link comments to specific products, designs, or vendors. Therefore, we could not draw any conclusions about specific products or the relative advantages of private-industry laboratory portals or portals tethered to organizational EHRs.

Third, only consumer self-reported data were gathered; our findings are based on subjective perceptions rather than directly observed user performance. Therefore, we may have overestimated the usability and understandability of web-based information—a respondent bias known as the “illusion of fluency” [32]. For example, participants were asked how well they understood the information, but we did not use a specific example and measure their understanding. Related work has shown that even experienced users of laboratory results portals can easily overlook abnormal values [27].

Finally, we only studied the perspectives of users; we did not gather the perspectives of nonusers. Therefore, all participants had experience using one or more laboratory results portals. This represents an important selection bias; users could be very different from nonusers in their personal goals, search strategies, technology literacy, and health literacy. Furthermore, we could not explore all potential deterrents to accessing these systems.

Conclusions

Through our questionnaire results, we identified the barriers and facilitators to using these systems and highlighted opportunities where such systems could be improved. We identified areas for improvement centered around usability, information, and displays. This study offers health care organizations and health information system developers general recommendations on how to better design these products to align with users’ needs and for optimal use.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Infographic of study findings.

[PDF File (Adobe PDF File), 263 KB - humanfactors_v10i1e42843_appl1.pdf ]

References


Abbreviations

EHR: electronic health record
NPS: Net Promoter Score

©Helen Monkman, Janessa Griffith, Leah MacDonald, Blake Lesselroth. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 12.06.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Designing and Developing an eHealth Program for Patients With Persistent Physical Symptoms: Usability Study

Oliver Rønn Christensen¹*, MSc; Leonora Hedegaard²*, MSc; Mette Trøllund Rask¹*, PhD; Jane Clemensen³*, PhD; Lisbeth Frostholm¹,⁴*, PhD; Marianne Rosendal¹,⁵*, MD, PhD

¹Research Clinic for Functional Disorders and Psychosomatics, Aarhus University Hospital, Aarhus N, Denmark
²Department of Culture and Learning, Aalborg University, Aalborg, Denmark
³Hans Christian Andersen Children’s Hospital, Odense University Hospital, Odense, Denmark
⁴Department of Clinical Medicine, Aarhus University, Aarhus, Denmark
⁵Research Unit for General Practice, Aarhus University, Aarhus, Denmark
*all authors contributed equally

Corresponding Author:
Oliver Rønn Christensen, MSc
Research Clinic for Functional Disorders and Psychosomatics
Aarhus University Hospital
Palle Juul Jensens Boulevard 11
Aarhus N, 8200 Denmark
Phone: 45 26540918
Email: oliver.ronn.christensen@gmail.com

Abstract

Background: Patients with persistent physical symptoms presenting in primary care are often affected by multiple symptoms and reduced functioning. The medical and societal costs of these patients are high, and there is a need for new interventions tailored to both the patients and health care system.

Objective: This study aimed to examine the usability of an unguided, self-help treatment program, “My Symptoms,” developed to assist patients and general practitioners in symptom management.

Methods: In all, 11 users (4 patients with persistent physical symptoms and 7 laypeople) participated in web-based thinking-aloud interviews involving the performance of predefined tasks in the program. Thematic analysis was used to categorize the severity of usability issues. General usability heuristics were cross-referenced with the usability issues.

Results: The analysis identified important usability issues related to functionality, navigation, and content. The study shows how therapeutic knowledge in some cases was lost in the translation of face-to-face therapy to a digital format. The user testing helped uncover how the functionality of the digital elements and general navigation of the program played a huge part in locating and accessing the needed treatment. Examples of redesign to mediate the therapeutic value in the digital format involving health care professionals, web developers, and users are provided. The study also highlights the differences of involving patients and laypeople in the interviews.

Conclusions: Taking the experience of common symptoms as a point of departure, patients and laypeople contributed to finding usability issues on program functionality, navigation, and content to improve the program and make the treatment more accessible to users.

(JMIR Hum Factors 2023;10:e42572) doi:10.2196/42572

KEYWORDS
eHealth; digital health; medically unexplained symptom; persistent physical symptom; self-management; usability; physical symptom; persistent symptom; unexplained symptom; symptom management; unguided; thinking aloud; think aloud
**Introduction**

**Background**

The experience of physical symptoms is a normal phenomenon. Most symptoms are self-limiting, but they may become persistent and lead to frequent contacts with health care providers. In general practice, 17% of patients are affected by persistent physical symptoms (PPS). These patients have an increased risk of disability and mental comorbidity impacting quality of life, health care use, and the ability to work [1,2].

Symptoms span a continuum from mild to severely disabling. This paper refers to symptoms that lead to repeated contact with general practice and may be associated with some degree of functional disability but do not reach the severity of a disease such as a functional disorder (Figure 1). The present Danish national treatment guidelines [3,4] recommend a stepped care model cohering to this continuum in which general practitioners (GPs) are expected to provide care for patients with mild to moderately severe symptoms, that is, PPS. However, specific treatment for PPS in general practice is almost nonexisting [5]. General practice is characterized by high workload and time-restricted consultations, and GPs may tend to focus on investigations to rule out severe disease, without providing guidance to patients on how to manage their symptoms when tests and investigations come out negative [6]. Thus, there is an urgent need to improve the treatment of patients with PPS in primary care to support change in symptom perception and illness behavior to reduce patients’ risk of becoming chronically disabled.

**Figure 1.** Illness spectrum from symptom to disorders (based on Rosendal et al [7]).

---

**The My Symptoms Program and Study**

According to previous studies, internet-based self-help interventions may contribute to symptom alleviation and improved quality of life [8,9]. To assist GPs in symptom management and to offer patients with PPS a new treatment option, we developed a novel eHealth program, “My Symptoms.” The program content is inspired by cognitive behavioral therapy. It provides psychoeducation on symptoms and modules on the impact of lifestyle, stress and strain, thoughts, feelings, values, and self-care. Throughout the modules, interactive tools to support behavior change are embedded. The patient can interact with modules on his or her own accord (Figure 2). The content of “My Symptoms” is presented in various forms such as text, pictures, figures, interactive elements, audio, and video. The program is prescribed by the GP but is unguided, that is, no health care professional (HCP) will assist the patient in the use of the program. The program is a responsive web application that is accessible from computers, tablets, and smartphones through a web browser.

The overall framework of making the “My Symptoms” program lent itself to ideas from the participatory design research paradigm within health care [10,11]. Here, emphasis was on the democratization of the development from different stakeholders and participants via iterative processes. The development of “My Symptoms” followed three phases (Figure 3): phase 1, identification of needs [6]; phase 2, design and development; and phase 3, feasibility study. This paper reports on the usability studies conducted in phase 2 (Figure 3). The results from this study informed the content and structure of the program used in the feasibility study.

The objectives of the present study were (1) to investigate the usability of the “My Symptoms” program with a specific focus on how to improve functionality and navigation and (2) to explore how users could help improve the intuitiveness and user-friendliness of the program.
Methods

Overview

From 2020 to 2021, the last part of the development phase 2 (Figure 3) commenced with an emphasis on the usability of the program. To examine user experience with navigation and functionality in the program, we conducted thinking-aloud sessions asking participants to speak aloud while completing various tasks [13,14].

The project group developing the self-help program consisted of HCPs (GPs, a psychiatrist, psychologists, and a public health scientist), anthropologists, techno-anthropologists, and web developers. The techno-anthropologists conducted the thinking-aloud sessions, whereas the whole group was involved in the processing of results.

Ethical Considerations

All participants were informed orally and in writing about the study, and all participants gave their consent to participate. The study was approved by the Danish Data Protection Agency (J. no. 1-16-02-16-19). The Danish Act on Research Ethics Review of Health Research Projects is not applicable to qualitative studies. Therefore, ethical approval was not required from the Committee on Health Research Ethics in the Central Denmark Region.

Participants

We included a convenience sample of primary care patients and laypeople. In all, 6 GPs identified and invited 4 patients aged 18-65 years with PPS. These potentially eligible patients were informed orally and in writing about the project by the GPs and gave their consent to be contacted by a researcher from the project team. A project member screened consenting patients according to the selection criteria and finally included or excluded patients (Textbox 1).
We aimed for the inclusion of approximately 12 patients, considering the number of 5-8 participants proposed by Nielsen [16] and Virzi [17] on finding most of the usability issues. Moreover, when investigating usability iteratively, the number of participants needed is adjusted continuously based on data saturation, that is, in our case, the number of medium and critical issues. However, the restrictions caused by the COVID-19 pandemic [18] challenged the recruitment process in general practice. To finalize the study within its time limits, we therefore chose to supplement the user inclusion with 7 laypeople recruited through personal networks. As bothersome symptoms are a general phenomenon [19], we expected laypeople to be able to relate to current or prior symptom experiences. The patient group consisted of 4 patients; 50% (n=2) were female, and ages ranged from 24-57 years. One patient had 2 rounds of testing and interviewing, whereas the remaining patients had 1. The laypeople group consisted of 7 individuals; 57% (n=4) were female, and ages ranged from 20-31 years. In all, 3 laypersons had 3 rounds of testing and interviewing, 2 had 2 rounds, and 2 had 1 round.

### Usability Investigation by Thinking Aloud

To investigate usability, we applied the thinking-aloud method. The aim of this method was to “capture” the users’ thoughts as they navigated the “My Symptoms” program to gain insight into how they experienced the program in the context of actual use and what they found easy or difficult to do or understand [20]. The project group translated these verbalized thoughts into specific changes that needed to be made in the program. Due to the restrictions caused by the COVID-19 pandemic, we conducted the thinking-aloud sessions on the web [18]. We used screen sharing that allowed for easy observation of how the user navigated the program and recorded the sessions for subsequent analysis. We chose to let the users’ symptoms guide their way through the program to approximate actual use case scenarios [21]. When investigating the functionality of the interactive exercises, predefined tasks were provided with scenarios based on interviews from a preceding study [6]. The users were told to imagine being referred by their GP, coming home with a flyer with instructions on how to access and use the program. In the first of 3 rounds of testing, all users carried out the same 8 tasks related to log-in, filling out questionnaires and exercises, finding information about one’s symptom(s), and using interactive behavior change tools. For example, the interviewer would ask the user to access information about the most bothersome symptom and observe their behavior. Sometimes when issues arose, the interviewer asked, “what did you expect would have happened?” but mostly kept quiet until the participant had completed a task to not interfere with the participant’s experience.

Rounds 2 and 3 focused more on testing predefined, specific elements in the program rather than core functionality. From the first round of testing, we observed that laypeople and patients interacted similarly to buttons, sliders, and other interactive web elements, which was why we included more laypeople than patients for these rounds. Immediately after thinking aloud, a follow-up interview [22,23] was conducted inquiring about the users’ experience of the program. The questions were related to overall experience, relevance for everyday life, and the use of the internet for behavior change. The thinking-aloud testing and follow-up interview lasted 45-90 minutes. A total of 20 sessions were conducted.

### Data Analysis

All audio and video recordings were transcribed. Based on transcripts and notes taken during the thinking-aloud sessions, we identified usability issues and rated the severity of these as minor, medium, or critical. Encountered bugs were also flagged. The GitHub platform (GitHub, Inc) [24] was used to relay and manage bugs and usability issues to the web developers. Using a thematic analysis approach [25], all usability issues rated as medium or critical were coded based on content. Subsequently, these codes were mapped to 2 predefined categories: navigation and functionality. After categorization, we cross-referenced the emerged categories with general usability heuristics [26]. The usability heuristics offer a set of guidelines curated over decades of designing systems and identifying usability problems. They are often used to inform design decisions by experts within the field of human-computer interaction.

### Results

#### Categories and Core Issues

The thinking-aloud sessions gave insight into functionality and navigation. Additionally, the sessions revealed potentially problematic phrases and wordings that hindered the usability

---

**Textbox 1. Selection criteria for patients.**

<table>
<thead>
<tr>
<th>Inclusion criteria</th>
<th>Exclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Age 18-65 years</td>
<td>1. Severe mental disorder</td>
</tr>
<tr>
<td>2. Affected by persistent physical symptoms according to their general practitioner</td>
<td>2. Sick leave for more than 8 consecutive weeks</td>
</tr>
<tr>
<td>3. “Somewhat bothered” by at least 4 of 25 symptoms (scoring ≥2 on a Likert scale with each symptom from 0 “Not bothered at all” to 4 “Bothered a lot” by the Bodily distress syndrome (BDS) checklist [15])</td>
<td></td>
</tr>
<tr>
<td>4. Speak and understand Danish</td>
<td></td>
</tr>
</tbody>
</table>

**Inclusion criteria**

1. Age 18-65 years
2. Affected by persistent physical symptoms according to their general practitioner
3. “Somewhat bothered” by at least 4 of 25 symptoms (scoring ≥2 on a Likert scale with each symptom from 0 “Not bothered at all” to 4 “Bothered a lot” by the Bodily distress syndrome (BDS) checklist [15])
4. Speak and understand Danish

**Exclusion criteria**

1. Severe mental disorder
2. Sick leave for more than 8 consecutive weeks

---

**Selection criteria for patients.**

1. Age 18-65 years
2. Affected by persistent physical symptoms according to their general practitioner
3. “Somewhat bothered” by at least 4 of 25 symptoms (scoring ≥2 on a Likert scale with each symptom from 0 “Not bothered at all” to 4 “Bothered a lot” by the Bodily distress syndrome (BDS) checklist [15])
4. Speak and understand Danish

**Inclusion criteria**

1. Age 18-65 years
2. Affected by persistent physical symptoms according to their general practitioner
3. “Somewhat bothered” by at least 4 of 25 symptoms (scoring ≥2 on a Likert scale with each symptom from 0 “Not bothered at all” to 4 “Bothered a lot” by the Bodily distress syndrome (BDS) checklist [15])
4. Speak and understand Danish

**Exclusion criteria**

1. Severe mental disorder
2. Sick leave for more than 8 consecutive weeks
of the program. Inductively, a third category concerning content emerged, leaving us with 3 overall categories of issues. In Table 1, the results are presented according to these categories: functionality, navigation, and content, using examples from the data and the issues identified. In the following sections, we elaborate on some of the issues found within each category. These issues are marked by a footnote.

Table 1. Overview of results with categories and core issues.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Core issues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translating therapy into digital functionality</td>
<td>• Missing information on how to perform exercises(^a)</td>
</tr>
<tr>
<td></td>
<td>• Interactive exercises often need surrounding text before being used sufficiently</td>
</tr>
<tr>
<td></td>
<td>• Using examples of changing habits as options and not suggestions</td>
</tr>
<tr>
<td></td>
<td>• Confusion about “Archive” and “Close” buttons when using core tools</td>
</tr>
<tr>
<td>Designing navigation</td>
<td>• Too many submodules and pages(^b)</td>
</tr>
<tr>
<td></td>
<td>• Missing references for getting back to already known content</td>
</tr>
<tr>
<td></td>
<td>• Missing navigation buttons at the bottom of pages(^a)</td>
</tr>
<tr>
<td></td>
<td>• Too much blank space leaving users to miss content</td>
</tr>
<tr>
<td>Content guides program use</td>
<td>• Text-heavy pages(^a)</td>
</tr>
<tr>
<td></td>
<td>• Using quotation marks on health information decreases users’ perceptions of program legitimacy(^a)</td>
</tr>
<tr>
<td></td>
<td>• Missing language directed at the user</td>
</tr>
</tbody>
</table>

\(^a\)These core issues are further elaborated below.

**Translating Therapy Into Digital Functionality**

In the development phase, the interactive elements and exercises of the “My Symptoms” program required close collaboration between web developers and the project group as common therapeutic tools were translated into digital equivalents. The usability test revealed that the intended purpose of the therapeutic content in some instances had been lost in this translation. General examples of core issues that suffered from translational issues were missing information on how to perform exercises, the need for additional text or content to explain exercise use, generic examples on changing habits were not translated by the user and used as is, and confusion on how to use core tools such as the “Goal Staircase.” Most of these issues concerned exercises that had been translated from a physical context into digital entities. Such translational problems were especially evident with the “Sorting of Values” \(^{[27]}\) exercise.

In the “Sorting of Values” exercise, the user was supposed to select statements about the life values most important to him or her, such as “I value a healthy diet,” “I look for challenges,” and “My goal is to live in harmony with nature.” The exercise was then to sort the statements into the columns of “agree” or “disagree.” Finally, the user was prompted to choose 3 to 5 of the agreed statements into a new column to identify the most important values. Figure 4 shows how the web developers and the HCPs had manifested the concept of the exercise.

In the “Sorting of Values” exercise, only 1 of 7 users was able to complete the exercise. Completing this exercise was critical since the rest of the module depended on the “results” gained by completing it. This 1 user was able to access a “hidden” column that would only appear when all the value-statements had been sorted into the 2 above columns. Only then the user was able to sort the 3 to 5 most important statements into the previously hidden column, and the program would store these for later use. Figure 5 shows what the exercise looks like when completed.

In the physical version of the exercise, the patients would hold a deck of cards in their hands that would give them tactile feedback on how many cards or statements were left to sort. With the example of the “Sorting of Values” exercise, we noticed that one reason for stopping the sorting of values into the first 2 columns was the absence of knowledge on how many statements have to be sorted and how long it would take. Thus, the users figured it would be okay having some of the statements sorted into the 2 columns and moved on to the next page. The heuristic “visibility of system status” \(^{[26]}\) reminds us that the user in general should be kept informed on the state of the program, for example, through feedback on the current progress of a specific exercise or task.

Based on the findings, the project group developed a solution where the third column was now shown all the time (Figure 6) to display the goal of the exercise from the beginning. Furthermore, statements were grouped into a “scroll box,” hinting at the number of statements and how many need to be sorted. Additionally, a reset button was added to allow for increased user control.

The “Sorting of Values” exercise was one example of several issues on not getting enough information at the right time to complete the exercises. Other issues were, for example, related to missing tactile information such as the “Sorting of Values” exercise and missing guidance from a therapist in the digital format.
Figure 4. The "Sorting of Values" exercise in the "My Symptoms" program. Statements such as "I seek challenges" and "My faith gives me strength" must be sorted in columns of "disagree" (left) or "agree" (right).

Figure 5. The "Sorting of Values" exercise showing the third column, "the most important values" (yellow), used to store and remember statements for later use in the module.
Designing Navigation

When investigating navigation, core issues were regarding too many pages and submodules, the placement of navigation buttons, navigating back to known content, and too much blank space. These navigational issues were especially frustrating to new users trying to find their way through the program.

An example of a major navigational issue identified was navigating the different levels of the program. Knowing what page the user was on in relation to the rest of the program proved difficult. This was important if the user wanted to repeat or go back to an exercise in a new sitting. The thinking-aloud sessions demonstrated how most people tried to remember the location of the exercises by recalling the modules and the title on subpages. As a module could hold up to 3 levels and 10 subpages, this strategy proved difficult. Furthermore, the subpages of the modules were navigated by the next and previous page buttons bound to the top of specific pages. This made the navigation buttons disappear when scrolling down.

To alleviate the users’ need for recalling the different locations of the exercises, a “sticky” navigation bar was developed. This was done in reference to, for example, the heuristics “recognition rather than recall” and “visibility of the system status” [26]. The new navigation bar indicated at which level and page the user was, and it was shown all the time at the bottom of the page. We also reduced the number of levels of all modules to 2. Moreover, a breadcrumb trail was made visible showing the full extent of the path.

Initially, the program was aimed at allowing the users to navigate freely according to their symptoms and to let them decide what content would be appropriate or helpful to explore. However, the usability tests revealed a need to help the users better navigate between the overall modules. They needed a better framework for their journey in the program. Therefore, we decided to let the first 3 modules open gradually before providing free access to all modules. When a new module opens, the users are prompted with a text message on their telephone.

Content Guides Program Use

Analysis of the thinking-aloud and interview data revealed that specific program content could potentially hinder adequate program use. Core issues in this category were related to pages being text heavy, the perception of program legitimacy, and the use of language directed at the user. These issues could in many instances be attributed to the lack of therapeutic assistance to mediate program use, thus compensating for it with more text and content. These issues were especially evident when comparing feedback from patients and laypeople. Patients were more focused on the framing of the content in the program than laypeople. For example, in content, quotation marks were used to stress some of the medical terms. Most patients responded poorly toward this usage and figured the terms were made up, making them question the legitimacy of the program and their own experience with their symptoms. This was not an issue with laypeople. Patients also spent more time investigating content as means of navigation and interaction, even though we explicitly stated that reading the content thoroughly was not necessary during the usability testing. Contrarily to patients, laypeople found it bothersome to read too much text as a guide for navigation and interaction. Instead, they preferred being guided by different design cues. Patients, however, found it necessary to read most of the text to make sure they were on the right path while being guided by design cues. Patients who experienced symptoms fatigue and lack of concentration were unable to adhere to the thinking-aloud tasks to the same degree.
as other users, which they attributed to the pages being too text heavy. With patients tending to read most texts thoroughly, the program may be experienced as being too bothersome, thus decreasing motivation for use.

In the follow-up interviews, we found that users related to the content in different ways. For example, when we asked participants, “Are there elements in the program you would like to return to and if so which ones?” (Table 2), laypeople spoke more in terms of what elements they would use, and patients spoke more in terms of how they would use them.

Table 2. Excerpts from the follow-up interviews (translated from Danish).

<table>
<thead>
<tr>
<th>Elements</th>
<th>Laypeople</th>
<th>Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>My feelings</td>
<td>“I think it is interesting to see how my feelings can affect my physical symptoms and vice versa. That is useful also in a general sense” (Layperson 5)</td>
<td>“I like the connections between my feelings and symptoms, but they require a lot of energy to make. I would need to find a calm space, like the woods, where I can be with my thoughts without too many distractions. Then it would just be myself and my iPad” (Patient 2)</td>
</tr>
<tr>
<td>My sleep</td>
<td>“Yeah, that with the sleep patterns, I think I could use. The sleep registration tool” (Layperson 7)</td>
<td>“I haven’t really tried those things before. It seems like a good idea, if you want to gradually go earlier to bed; go to bed one hour earlier every day using the Goal Staircase. But, I’d might just use some paper to keep track of my sleep” (Patient 3)</td>
</tr>
</tbody>
</table>

Discussion

Principal Findings

The purpose of this study was to investigate the usability of the “My Symptoms” program by exploring functionality and navigation from a user perspective. Investigating the functionality of the program mostly revealed issues related to the translation of face-to-face therapeutic material into digital exercises. This was especially in relation to issues regarding the lack of transparency on how to complete exercises and lack of continuous feedback on exercise progression. With regard to navigation, most of the usability issues were about the number of subpage levels and lack of markers when users wished to go back to exercises or content. By reducing the number of levels of subpages and using different design cues, we sought to evoke the recognition of symbols rather than recall of page titles.

The content of the program was not the target of our investigation but became a stepping stone for understanding how content also guides navigation and functionality. This was especially evident from the patients’ feedback. Although, differences in the inclusion of patients and laypeople were found, including laypeople was helpful when there was a lack of patients or patients were challenged cognitively—especially since the experience of bothersome physical symptoms is a common phenomenon. Laypeople and patients draw on the same type of IT-related schemata, for example, knowledge of browsing the web, using social media, and more. Moreover, they also draw on the experience on common symptoms.

Discussion of Results

Designing behavior change exercises requires a lot of attention to the communication between the HCPs and web developers. The process of improving the functionality of the “Sorting of Values” exercise was attributed to the iterative process of continuously testing the exercise with users. Testing the exercise multiple times helped us make a digital therapy that makes sense at all levels: therapeutic model, technical capacity, and user-friendliness. Not testing the usability would have made the exercise, and the rest of the “My Values” module, inaccessible to most users. Here, a pragmatic approach of understanding both the therapeutic models and the system capability was necessary. For example, the iterative process allowed different versions of the same exercise to be investigated multiple times by different participants, making the users, web developers, and HCPs all a part of the designing process [12,28].

The user feedback also helped inform uncertainties within the project team on deciding specific design solutions. Furthermore, a systematic review on user involvement in the development of patient decision aids found that projects could be more iterative and that reporting on the differences in design changes between iterations could help explain the rationale behind the finale product [29]. Here, we used the “Sorting of Values” exercise as an example of what rationales went into developing it and how we came about it.

In this study, laypeople acted as surrogates for patients. Knowing when and how to include laypeople is valuable when there is a lack of the intended end user. The use of surrogates can be beneficial in getting rid of the most critical usability issues before gaining access to a limited end user group [30]. When using surrogates, one must consider to which degree the real end user and the surrogate share the same characteristics [31].

In this study, the focus on common symptoms created a general foundation for testing the program for both patients and laypeople. By creating task scenarios based upon the real end users’ needs found in a preceding study [6] and comparing the real end users who are available to the surrogates, we were able to come up with and iteratively tailor the usability investigation in favor of our real end users. Using laypeople and patients in testing the usability of the “My Symptoms” program thus offered great insights to the general navigation and functionality. With patients experiencing, for example, frequent headaches, some lower back pain, and others fatigue, there was a widespread variation of how the patients’ symptoms manifested into in-program behaviors.

Strengths and Limitations

Through the thinking-aloud tests, we observed user behavior, and through the interviews, we obtained comments and suggestions for design changes. These data were cross-referenced with Nielsen’s heuristics [26]. Using the heuristics in combination with user-generated data, we wanted...
to alleviate the concern of bias from letting design changes being solely driven by experts’ interpretation of the heuristics [32], which we consider to be a strength.

Although the web-based tests and interviews made for easy observation and recording, it limited our contextual understanding of the practical setup and the environment of the users in which they might use the program. Likewise, our aim for consistency in testing solely with a desktop setup did not necessarily match our patients’ use cases, pointing toward further investigations into the use of, for example, tablets and smartphones, when using the “My Symptoms” program. Thus, future research on improving the “My Symptoms” program and other studies alike could benefit from focusing on contextual inquires [33] with patients both via observations and self-reported data that are, for example, enabled by different kinds of cultural probes [34].

In usability engineering, it is well known that by understanding the user’s mental model, or schemata, we are able to come closer to a conceptual model [35] by which we can come up with suggestions for changing the program. Moreover, the interviews with patients and observations in general practice preceding this study [6] helped inform the patient use cases when making the task scenarios for the thinking-aloud investigation. This process was similar to the concept of creating personas [36,37]. This also helped us make more realistic scenarios for including laypeople. However, even though laypeople were able to provide useful knowledge on their experience in the program—navigating out from their mental model [35] in a way similar to patients—laypeople cannot provide us with the experience of testing the program as someone experiencing persistent symptoms. As the patients pointed out, pages may be too text heavy; thus, a greater focus on health literacy is warranted. Therefore, although bodily sensations and symptoms are part of human life, recall bias on part of laypeople may exist, pointing to a limitation of this study.

Although the small sample size could be seen as a limiting factor, using the thinking-aloud method to highlight usability problems of immediate use is known to be an effective tool when having a small sample size. Because of the amount of detailed data the method provides, only 5-8 users are necessary to detect 80% to 85% of the usability problems [16,17]. Furthermore, according to a scoping review [38], different open-ended qualitative investigations should be deployed instead of a single method [39-41]. Second, immediate use should be explored rather than only retrospective investigations, such as interviews [42]. Third, during the development stage, it may be more beneficial to be informed by in-depth investigations using multiple cycles of exploration [28] and tests with a smaller group of participants than using a larger group of participants only once [38].

Interviewing the users before and after the thinking-aloud test enabled us to understand to which degree the users were accustomed to using web applications and IT in general. The interviews also gave us a chance to know how the patients were bothered by symptoms, helping us understand how the symptoms might affect the patients’ in-program behavior. Furthermore, the interviews helped us dive into issues that occurred during the thinking-aloud test, giving the users a way to suggest design changes retrospectively.

**Conclusions**

Creating a digital self-help treatment program demands special attention to user-friendliness and intuitiveness. Program usability can make the difference between getting the needed treatment or not. User feedback helped improve the usability of the program and revealed how therapy sometimes was lost in the digital translation. This was especially relevant in relation to the themes: functionality, navigation, and content. Here, reducing the number of subpages, providing users feedback on tasks, and being sensitive to the framing of the content increased user satisfaction. Using the thinking-aloud method and heuristics in combination enabled us to upscale the specific design iterations into more broadly defined design statements that are applicable throughout the program. Furthermore, the usability testing helped facilitate knowledge sharing between different professions as a precondition for a successful program development facilitated by an iterative development process.

**Acknowledgments**

The project group wants to acknowledge the efforts of the general practitioners who assisted with patient recruitment. We also want to acknowledge the participating patients and laypeople who spend many hours testing and helping improve the program. Lastly, we thank the web developers for making the programming and design.

This work was supported by the Innovation Fund Denmark (8056-00040B), TrygFonden (7549-07), and the Quality Improvement Committee for General Practice in the Region of Southern Denmark (KEU-Syd; 04/18).

**Conflicts of Interest**

None declared.

**References**


2. Rask MT, Rosendal M, Fenger-Grøn M, Bro F, Ørnøl E, Fink P. Sick leave and work disability in primary care patients with recent-onset multiple medically unexplained symptoms and persistent somatoform disorders: a 10-year follow-up of


Abbreviations

GP: general practitioner
HCP: health care professional
PPS: persistent physical symptoms

Please cite as:
Christensen OR, Hedegaard L, Rask MT, Clemensen J, Frostholm L, Rosendal M
Designing and Developing an eHealth Program for Patients With Persistent Physical Symptoms: Usability Study
JMIR Hum Factors 2023;10:e42572
URL: https://humanfactors.jmir.org/2023/1/e42572
doi: 10.2196/42572
PMID:36753312
Evaluation of an Electronic Care and Rehabilitation Planning Tool With Stroke Survivors With Aphasia: Usability Study

Nadia Davoody¹, MSc, PhD; Aboozar Eghdam¹, MSc, PhD; Sabine Koch¹, MSc, PhD, FACMI, FIAHSI; Maria Hägglund¹, MSc, PhD

Health Informatics Centre, Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Stockholm, Sweden

Corresponding Author:
Nadia Davoody, MSc, PhD
Health Informatics Centre, Department of Learning, Informatics, Management and Ethics
Karolinska Institutet
Tomtebodavägen 18 A
Stockholm,
Sweden
Phone: 46 (0)8 524 864 86
Email: nadia.davoody@ki.se

Abstract

Background: Patients with chronic illnesses with physical and cognitive disabilities, particularly stroke survivors with aphasia, are often not involved in design and evaluation processes. As a consequence, existing eHealth services often do not meet the needs of this group of patients, which has resulted in a digital divide.

Objective: The aim of this study was to examine the effectiveness and user satisfaction of an electronic care and rehabilitation planning tool from the perspective of stroke survivors with aphasia. This would help us gain knowledge on how such a tool would need to be adapted for these patients for further development.

Methods: Usability tests were conducted with 9 postdischarge stroke survivors with aphasia. Effectiveness was measured using task-based tests, and user satisfaction was studied through qualitative interviews at the end of each test. All tests were audio recorded, and each test lasted approximately 1 hour. The data were analyzed using qualitative content analysis. As the tool can be used by stroke survivors either independently or with some support from their next of kin or care professionals, the research group decided to divide the participants into 2 groups. Group 1 did not receive any support during the tests, and group 2 received some minor support from the moderator.

Results: The results showed that the care and rehabilitation planning tool was not effective for stroke survivors with aphasia, as many participants in group 1 did not accomplish the tasks successfully. Despite several usability problems and challenges in using the tool because of patients’ disabilities, the participants were positive toward using the tool and found it useful for their care and rehabilitation journey.

Conclusions: There is a need to involve patients with chronic illnesses more in the design and evaluation processes of health information systems and eHealth services. eHealth services and health information systems designed for this group of patients should be more adaptable and flexible to provide them with appropriate functionalities and features, meet their needs, and be useful and easy to use. In addition, the design and evaluation processes should be adapted, considering the challenges of this patient group.

(JMIR Hum Factors 2023;10:e43861) doi:10.2196/43861

KEYWORDS
usability testing; stroke; aphasia; eHealth; rehabilitation; co-design; evaluation; user-centered design; effectiveness; user satisfaction; mobile phone
**Introduction**

**Background**

Health information systems (HISs) and eHealth services have become essential parts of today's health care. For a long time, the use of information and communications technology (ICT) in health care has been increasing. Sensors, electronic health records, home monitoring for older adults, wearables, and different assistive technologies have been designed for health care professionals, patients, and their next of kin. Telemedicine and the design of assistive tools for cancer care [1], cardiovascular diseases [2,3], and older adult care at home [4] are some examples in this area. Traditionally, patients and their next of kin have received eHealth services without being active in the development process. Nevertheless, their input in the design of the tools and services has become a crucial part of this context. The design techniques in health care have moved from traditional system design methods to interactive design methods with the users involved and focus on the interaction between them and eHealth services. In recent years, users have acted as partners in the development process, and co-design methods have become more popular in health care. Even though implementing this type of design in health care is challenging, the benefits are enormous. There is a growing interest in user involvement in the development of electronic health interventions, and patient participation has become an inseparable part of the design process [5]. Despite the fact that several attempts have been made to design HISs and services for patients [6], there is still a need for appropriate ones for different patient groups. Involving the end users is crucial in a user-centered design process [7,8] as it improves the understanding of the users' needs and task requirements. It also plays an essential role in enhancing the iterations of the design and evaluation throughout the design process. Studies have shown that co-design improves patient knowledge, increases patient satisfaction, and improves care experience and sense of participation [9-12]. Several studies have focused on involving users and increasing their level of involvement in recent years [13-15]. Despite the growing popularity of involving patients with chronic illnesses with, for example, heart diseases, stroke, and obstructive pulmonary diseases in the design process [16-19], the number of eHealth services designed together with patients with either physical or cognitive disabilities or both is still low. This has resulted in a digital divide in which patients with the greatest needs do not have access to appropriate services [20,21]. Using a user-centered design and involving the end users could be a solution for reducing the digital divide for patients and citizens who are in need of accessing appropriate eHealth services to improve their care and rehabilitation processes.

Although ICT provides opportunities to improve, for example, the quality of health care [22,23], it also has its weaknesses and may affect health care and patients negatively [24]. Therefore, evaluation of HISs and eHealth services is recommended [24,25]. The formative evaluation of HISs and eHealth services is a crucial part of the user-centered design process. Recently, patients have been involved in several evaluation studies [26-30]. It is of great importance to design and evaluate HISs and services with and for patients to enable them, independently or with some support from their next of kin or health care professionals, to access and use the appropriate ICT tools that meet their needs. Studies involving older people with chronic diseases such as dementia have also shown the importance of the involvement of end users in the development of interventions as the usability results show an overall satisfaction with the platform [31]. The Medical Research Council guidelines for developing and evaluating complex health care interventions also stress the importance of formative evaluations and feasibility studies [32,33].

The clinical context of this study is stroke care. Stroke is one of the major global health problems causing death and adult long-term disabilities [34]. Stroke survivors often have several physical and cognitive disabilities and require care and rehabilitation from different care providers. Although they have long-term needs for support and rehabilitation and could potentially greatly benefit from the use of eHealth, this patient group is rarely involved in the design and evaluation of ICT tools. In addition, assistive technologies are often complex for patients with cognitive disabilities, and there are limited ones designed specifically for this group of patients [35]. A specific condition that many stroke survivors experience is aphasia, which results in difficulties in verbal expression, reading or writing, and understanding what others are saying [36,37]. Stroke survivors with aphasia are even less involved in the design and evaluation of different eHealth services. In the area of aphasia, there have been some studies focusing on the rehabilitation and speech-language treatment of patients with aphasia using ICT that show promising results [38,39]. In addition, a study focused on the key design features that can enhance the accessibility of mobile technology for people with aphasia [40].

In this study, we aimed to evaluate a care and rehabilitation planning tool, *My care plan*, that was designed and developed together with postdischarge stroke survivors (N=12) in our previous studies [41-43]. However, the patients participating in the design process were stroke survivors who did not have aphasia, and in this study, we aimed to explore if the tool could also be used by stroke survivors with aphasia or if it would need to be redesigned to meet their specific needs. Therefore, the participants in this study were patients who had a diagnosis of aphasia and were registered in a course for patients with aphasia in an education center in Stockholm, Sweden.

**My Care Plan**

In previous studies [41-43], we designed an electronic care and rehabilitation planning tool. The development was conducted according to a user-centered design approach involving 12 postdischarge stroke survivors with milder physical and cognitive impairments. They were living at home and could handle computers. However, none of them had aphasia. The idea was that postdischarge stroke survivors, independently or with some support from their next of kin, should be able to use the tool throughout their care and rehabilitation processes at home. The tool consists of 2 main parts, namely, *My rehabilitation and Administrative and health related information*.
My rehabilitation mainly focuses on establishing a rehabilitation plan by identifying problems and planning goals and activities (Figure 1). Currently, a neurology team consisting of a speech therapist, a counselor, a physiotherapist, and an occupational therapist visits a postdischarge patient of stroke and establishes a paper-based rehabilitation plan together with the patient and possibly their next of kin. The team, together with the patient, identifies the patient’s problems and defines and documents the intended goals and activities [44]. The idea with the electronic care and rehabilitation planning tool was that patients be able to access necessary information during their journey along with the establishment of a rehabilitation plan either independently or together with their next of kin or care professionals in a neurology team. Through the care and rehabilitation planning tool, the patients are able to document their problems and define their goals and related activities. As rehabilitation is a crucial part of the recovery process after a stroke, there is a need for a specific and clear goal-setting process. Therefore, we consciously decided to provide postdischarge patients with 2 different types of goals, namely, “simple” (eg, being able to talk to others) and Specific, Measurable, Achievable, Relevant, and Time-bound (SMART; eg, being able to read a book within 1 month) goals. Currently, the care professionals in neurology teams in Stockholm County work with almost the same types of goals and measure the patients’ progress using different scales depending on the patients’ problems and activities. However, in our care and rehabilitation planning tool, we used goal attainment scaling (GAS) [45] to quantify the achievement of the patients’ predefined goals. The expended goals, their importance, the difficulty level, the expected outcomes, and the baseline for the patient’s condition before the training are essential for using the GAS methods. Patients can independently or together with their next of kin or care professionals in a neurology team use a 5-point scale to obtain an overview of their achievements. The other part of the tool consists of administrative and health-related information, such as my calendar, my notes, my medication, my disabilities, my care contacts, reminders, my rights and responsibilities, and my assistive tools.

Postdischarge stroke survivors in the Stockholm County Council were involved throughout the requirement analysis and design process. The stroke survivors had milder physical and cognitive disabilities, were living at home, and could handle computers. The tool was designed using a user-centered design [46] approach and was developed based on the patients’ information needs throughout their care and rehabilitation processes. The design process started with interviews with health care professionals and focus groups with stroke survivors without aphasia. The patients’ information needs were collected, and appropriate eHealth services were identified. Paper-based prototypes were then designed together with postdischarge stroke survivors and discussed in further focus groups. On the basis of the feedback from participants, an electronic care and rehabilitation plan was then developed, and its preliminary version was evaluated with stroke survivors other than those who were involved in the design process. The electronic care and rehabilitation plan was then improved based on the feedback from the preliminary evaluation and additional focus groups with other postdischarge stroke survivors [43]. The final version was then evaluated from the care professionals’ perspective using the Unified Theory of Acceptance and Use of Technology [47]. Figure 2 illustrates the process of design and evaluation of the care and rehabilitation plan. In this study, we focused on evaluating the latest version of the prototype with a number of postdischarge stroke survivors with aphasia.

Figure 1. The home page and the overview page of the care and rehabilitation planning tool.
Methods

Overview
The electronic care and rehabilitation plan was evaluated through a number of usability tests. A usability test is a technique used in user-centered design for evaluating a product or service by testing it with representative end users [48]. The usability tests in this study were performed in April 2016 and will be described in the following sections, along with the participants and their recruitment.

Description of the Usability Tests
A usability testing plan including 10 tasks focusing on finding information and establishing a rehabilitation plan by defining, for example, problems, goals, and activities was designed. In total, 3 pilot tests were performed with 1 usability expert, 1 patient of stroke, and 1 next of kin of a patient of stroke to validate the test and tasks.

Effectiveness and user satisfaction were the focus of this study according to the International Organization for Standardization 9241-11 guidance [49]. Effectiveness in this study focused on finding information and establishing a rehabilitation plan, including defining the problems, goals, and activities. Therefore, we explored the effectiveness of creating such a rehabilitation plan and finding necessary information in the care and rehabilitation planning tool. With regard to user satisfaction, we aimed to study user expectations, opinions, and preferences. We did not measure efficiency as participating patients had several cognitive and physical disabilities. In addition, the participants carried out the tasks while continuously thinking out loud throughout the whole task-based performance. This made the measurement of efficiency quite impossible. In total, 30 minutes were dedicated to task performance in each test. We used qualitative interviews instead of the System Usability Scale to measure user satisfaction as we were dealing with a group that had difficulties understanding what they read. Asking the participants to answer 10 questions with a 5-point scale rating would have been quite impossible as they had poststroke fatigue and several cognitive and physical disabilities.

Information letters were provided to all participants, and consent forms were obtained during the test sessions. A total of 4 tests were performed at Karolinska Institutet, and 5 tests were carried out in another location in Stockholm with participants who had difficulties getting to Karolinska Institutet. All sessions were audio recorded, and screen activities were video recorded using Camtasia Studio 8 software (version 8.4, TechSmith) for retrospective analysis by the research group. The computer used in this study was a Dell laptop running Windows 7 (Microsoft Corporation). The qualitative material obtained from the tests was transcribed verbatim and analyzed using an inductive content analysis approach according to Graneheim and Lundman [50]. All the authors were involved in creating the test tasks. All the tests were performed by the first author (ND). The first author (ND) went through the participants’ responses and conducted the preliminary data analysis. She grouped the responses into different categories. To ensure the credibility and trustworthiness of the results, the codes and categories were then discussed with other authors (AE, SK, and MH), and necessary changes were made. The process continued iteratively until a consensus on the categories and subcategories was reached. All authors were engaged in deciding the final codes and categories in this study. Prolonged engagement with the data, member checks, and the team’s unique combination of researchers with experience in qualitative research and evaluation studies allowed us to discuss and identify accurate codes and categories in this study. Quotations were extracted from the interview transcripts to illustrate core categories.

Each test took approximately 1 hour, and during each session, a short description of the study was given to the participants. Each test session was divided into 3 parts. The first part included an introduction to the test and receiving the consent form and the demographic questionnaire filled out by each participant. The consent form and introduction to the study and test were read by the moderator at the beginning of each test. The second part of the test included the tasks provided separately on paper and read aloud by the participants. During the last part of the test, participants were asked to answer some questions about different parts of the tool with a focus on user satisfaction and fill in a multiple-choice questionnaire based on resources from the work by Rubin and Chisnell [48] about the usefulness of the tool. The multiple-choice questionnaire consisted of 4 questions in which the participants, for example, expressed why it was necessary for them to have access to the care and rehabilitation planning tool and how useful it would be in their daily life. The test procedure for each participant is presented in Table 1.
Table 1. The procedure for the usability test for each participant.

<table>
<thead>
<tr>
<th>Step</th>
<th>Activity</th>
<th>Description</th>
<th>Measurement</th>
<th>Instrument</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td>Consent form</td>
<td>Patients approved their participation in the study</td>
<td>N/A&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Paper form (read by the moderator)</td>
</tr>
<tr>
<td>Step 2</td>
<td>Questionnaire (demographic information)</td>
<td>Sex, age, time since stroke occurrence, experience of having an electronic or paper-based rehabilitation plan, experience of using ICT&lt;sup&gt;b&lt;/sup&gt;, an overview of disabilities, and background (technical or nontechnical)</td>
<td>N/A</td>
<td>Paper form (filled with the help of the moderator)</td>
</tr>
<tr>
<td>Step 3</td>
<td>Introduction to the study and the test</td>
<td>Short description of the overall aim of the study and an introduction to the usability test</td>
<td>N/A</td>
<td>Paper form (read by the moderator)</td>
</tr>
<tr>
<td>Step 4</td>
<td>Introduction to the tool</td>
<td>Short description of the tool’s functionality</td>
<td>N/A</td>
<td>Video (short video of a screen recording with audio instructions)</td>
</tr>
<tr>
<td>Step 5</td>
<td>Test tasks</td>
<td>Part 1: finding information; part 2: establishing a rehabilitation plan</td>
<td>Performance effectiveness and completion of the tasks</td>
<td>Camtasia (TechSmith)</td>
</tr>
<tr>
<td>Step 6</td>
<td>Posttest interview</td>
<td>Questions on user expectations, opinions, and preferences</td>
<td>User satisfaction and open-ended questions</td>
<td>Paper form (read by the moderator)</td>
</tr>
<tr>
<td>Step 7</td>
<td>Posttest questionnaire</td>
<td>Questions about the usefulness of the tool</td>
<td>Multiple-choice questions</td>
<td>Paper form (read by the moderator and filled out by the participants)</td>
</tr>
</tbody>
</table>

<sup>a</sup>N/A: not applicable.<br><sup>b</sup>ICT: information and communications technology.

**Study Participants**

As the idea behind the design of the care and rehabilitation plan was that stroke survivors could use it independently or with some support from their next of kin or health care professionals, we decided to divide the participants into 2 groups (group 1 and group 2) to compare their performance in accomplishing the tasks. The participants were divided into the two groups based on the difficulties they faced and the level of support they received during the tests. Group 1 included patients who did not need any support and could follow the steps and accomplish some of the tasks successfully. In contrast, group 2 included patients that faced many challenges in performing the tasks and, therefore, received minor support during the test. In total, 6 usability tests of the My care plan tool were performed with 67% (6/9) of the postdischarge stroke survivors with aphasia (group 1). In addition, 3 usability tests were performed with the other 33% (3/9) of the postdischarge stroke survivors with aphasia (group 2). The patients in group 2 received some minor support from the moderator for performing the tasks. The support was mainly focused on asking questions such as “Do you see any rehabilitation plan?” or “Is that icon about assisting tool?”

The inclusion criterion for the participants in this study was that the time of stroke occurrence should not have been >5 years before. The participants had to live at their homes, have milder cognitive and physical disabilities, be able to talk with only minor difficulties, have aphasia, and be able to handle computers. Tables 2 and 3 provide an overview of the demographic information about the participants in this study and their experiences of, for example, having a rehabilitation plan or using different types of technology. The study was performed with patients with a diagnosis of aphasia from an education center in Stockholm.
Table 2. Participants’ characteristics (group 1), their experience of having rehabilitation and a rehabilitation plan and using technical devices, and a list of their disabilities.

<table>
<thead>
<tr>
<th>Participant ID</th>
<th>Sex</th>
<th>Age range (years)</th>
<th>Time since stroke occurrence</th>
<th>Experience having an electronic or paper-based rehabilitation plan</th>
<th>Still receiving rehabilitation from neurology teams</th>
<th>Experience using technology</th>
<th>Current living situation</th>
<th>Cognitive or physical disabilities</th>
<th>Technical or nontechnical background</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant A</td>
<td>Female</td>
<td>51-60</td>
<td>1-2 years</td>
<td>Paper-based rehabilitation plan</td>
<td>No</td>
<td>Computer and iPad</td>
<td>Partner</td>
<td>Speaking and writing difficulties and concentration problems</td>
<td>No</td>
</tr>
<tr>
<td>Participant B</td>
<td>Female</td>
<td>71-80</td>
<td>4.5 years</td>
<td>Paper-based rehabilitation plan</td>
<td>No</td>
<td>Computer</td>
<td>Alone</td>
<td>Speaking and writing difficulties</td>
<td>No</td>
</tr>
<tr>
<td>Participant C</td>
<td>Female</td>
<td>51-60</td>
<td>3 years</td>
<td>Paper-based rehabilitation plan</td>
<td>No</td>
<td>Computer and iPad</td>
<td>Partner</td>
<td>Right body side weakened</td>
<td>No</td>
</tr>
<tr>
<td>Participant D</td>
<td>Male</td>
<td>51-60</td>
<td>3 years</td>
<td>Paper-based rehabilitation plan</td>
<td>Yes</td>
<td>Computer, smartphone, and iPad</td>
<td>Partner, children, and other next of kin</td>
<td>Half-sided paralysis and speaking difficulties</td>
<td>No</td>
</tr>
<tr>
<td>Participant E</td>
<td>Male</td>
<td>61-70</td>
<td>2 years and 3 months</td>
<td>Paper-based rehabilitation plan</td>
<td>No</td>
<td>Computer, smartphone, and iPad</td>
<td>Partner</td>
<td>Aphasia</td>
<td>Yes</td>
</tr>
<tr>
<td>Participant F</td>
<td>Male</td>
<td>51-60</td>
<td>2 years and 9 months</td>
<td>No</td>
<td>No and had never received rehabilitation from neurology team</td>
<td>Computers</td>
<td>Alone</td>
<td>Aphasia, writing problems, poor eyesight, and problem with right side of body</td>
<td>No</td>
</tr>
</tbody>
</table>

Table 3. Participants’ characteristics (group 2), their experience of having rehabilitation and a rehabilitation plan and using technical devices, and a list of their disabilities.

<table>
<thead>
<tr>
<th>Participant ID</th>
<th>Sex</th>
<th>Age range (years)</th>
<th>Time since stroke occurrence</th>
<th>Experience having an electronic or paper-based rehabilitation plan</th>
<th>Still receiving rehabilitation from neurology teams</th>
<th>Experience using technology</th>
<th>Current living situation</th>
<th>Cognitive or physical disabilities</th>
<th>Technical or nontechnical background</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant G</td>
<td>Male</td>
<td>41-50</td>
<td>2 years and 7 months</td>
<td>Oral rehabilitation plan</td>
<td>Yes</td>
<td>Smartphones</td>
<td>Alone</td>
<td>Half-sided paralysis</td>
<td>No</td>
</tr>
<tr>
<td>Participant H</td>
<td>Male</td>
<td>51-60</td>
<td>1-2 years</td>
<td>Paper-based rehabilitation plan</td>
<td>Yes</td>
<td>Computers</td>
<td>Partner</td>
<td>Speech difficulties, walking problems, and right arm and leg not fully functional</td>
<td>No</td>
</tr>
<tr>
<td>Participant I</td>
<td>Male</td>
<td>41-50</td>
<td>1-2 years</td>
<td>Paper-based rehabilitation plan</td>
<td>No</td>
<td>Computer</td>
<td>Alone</td>
<td>Writing and speech difficulties and balance and memory problems</td>
<td>No</td>
</tr>
</tbody>
</table>

Overview of the Tasks

The effectiveness of the tool was studied for both groups of participants, namely, the 67% (6/9) of the participants who did not receive any support during the tests and the 33% (3/9) of the participants who received minor support for performing the tasks. A list of the tasks is shown in Textbox 1.
Textbox 1. An overview of the tasks in the study.

- Task 1: find out which primary care center you are listed at.
- Task 2: find out if you have any problem in your rehabilitation plan.
- Task 3: add a new problem regarding your walking. Use the predefined International Classification of Functioning, Disability and Health (ICF) codes. Link it to occupational therapy.
- Task 4: add a new problem for your stress without using ICF codes. Link it to counseling.
- Task 5: add a simple personal goal. Add a description and link it to your stress problem.
- Task 6: add a new Specific, Measurable, Achievable, Relevant, and Time-bound (SMART) goal and link it to your walking problem. Choose "as bad as it can be" in step 3/4 and then choose “difficult” and “very important” in step 4/4.
- Task 7: decide on a new review point for your SMART goal for October 31. Choose Klocka 1 PM.
- Task 8: add a new activity for your SMART goal. The activity should start on November 27 Klocka 3 PM and be repeated every Thursday until December 18.
- Task 10: do the review for your SMART goal. Choose “Mycket bättre.”

Ethics Approval and Consent to Participate

Ethics approval for the study was obtained from the Regional Ethics Committee of Stockholm on January 19, 2012 (2011/2093-31/5). Information letters were provided to all participants, and consent forms were obtained during the test sessions.

Results

Task Performance by Participants in Groups 1 and 2

The results of task performance by both groups are presented in the following sections.

Table 4. An overview of the task performance of participants in group 1. The blank cells to show that some participants did not succeed in completing some of the tasks.

<table>
<thead>
<tr>
<th>Tasks</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant A</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participant B</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participant C</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participant D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Participant E</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participant F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>

Participants With No Support During the Tests (Group 1)

The analysis of the results from group 1 showed that only 17% (1/6) of the participants were able to complete half of the tasks successfully. The other 83% (5/6) of the participants had major difficulties performing the tasks, especially those related to the establishment of a rehabilitation plan. These tasks required practice, concentration, and several mouse clicks to be implemented. In total, 83% (5/6) of the participants were able to perform at least one of the tasks related to information seeking. A total of 17% (1/6) of the participants were not able to perform any of the 10 tasks. Table 4 provides an overview of task performance of the participants in group 1.

Participants With Some Minor Support During the Tests (Group 2)

Task performance was different for the second group of participants. All participants (3/3, 100%) performed the first 2 tasks on finding information without any support from the moderator. Generally, participants in both groups accomplished the information-seeking tasks more successfully than the rehabilitation tasks. This might be due to the large icons for different content, which made the necessary information visible and easy to access for the users. In group 2, most of the tasks related to the establishment of a rehabilitation plan (adding problems, goals, activities, mood history, and assessment of goal achievement) were accomplished successfully with some minor support and cues from the moderator (eg, Is there any button you can push to add a new goal?). Table 5 provides an overview of the task performance of the participants in group 2.
Table 5. An overview of the task performance of the participants in group 2.

<table>
<thead>
<tr>
<th>Tasks</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant G</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Participant H</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Participant I</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

The results of task performance in both patient groups showed that the tool is not effective for postdischarge stroke survivors with aphasia as most of the participants (6/9, 66%) were not able to accomplish most of the tasks. However, the results from group 2 show that all participants (3/3, 100%) were able to accomplish most of the tasks successfully with some minor support from the moderator.

Qualitative Analysis of the Users’ Feedback

Overview

The content analysis of the posttest interview resulted in the 4 categories presented in following sections. The results in these sections are from participants in both groups. Textbox 2 provides an overview of the categories and subcategories identified in this study.

Textbox 2. An overview of the themes and categories.

- Satisfaction
  - Time to learn and support from next of kin
  - The appropriate time for using the tool after stroke occurrence
- Design implications related to fatigue and concentration difficulties
  - Information overload
  - Complexity of concepts
- Visualization
  - Using graphics
  - Results visualization
  - Using color-coding
- Perceived usefulness

Satisfaction

Overview

Despite the fact that the tool was not effective for postdischarge stroke survivors with aphasia, the posttest interview showed that the participants were positive toward using the electronic care and rehabilitation plan as a supporting tool for tracking their goals and activities. All participants except 1 (8/9, 89%) believed that they could or wanted to use the tool:

Yes, absolutely, I would like to [use the tool]. I use a computer every day, so I think it’s enjoyable [to use the tool]...It is fun to use the electronic care and rehabilitation plan. I think it is fun. [Participant A; group 1]

Of course, it is nice [to use the tool]. There is a lot. There are things like this [information in the tool]. It is great. [Participant B; group 1]

Sure, it was really difficult to understand but I do not know...I can use it. [Participant C; group 1]

It would be great; it is easy to access things here. [Participant F; group 1]

Some participants (3/9, 33%) believed that the tool was not difficult to use. However, most of the participants (6/9, 66%) found some parts of the tool quite complicated, for example, establishing a rehabilitation plan by adding goals and activities. A participant in group 1 believed that all parts of the tool were difficult, complex, and not easy to understand:

It was not so simple. It is a new program. You must learn it, it is pretty easy to click on [icons] and then go back and do it again. [Participant A; group 1]

I think you could make it much easier for us [stroke survivors]. It was very difficult for me to choose between all things in each page...It was very difficult for me to choose between goals and activities. Very difficult to understand. [Participant C; group 1]

Time to Learn and Support From Next of Kin

One of the participants mentioned that they could handle the tool if there was enough time to spend on working with the tool and, if necessary, they received some support from their next of kin:

I think I can [handle the tool]. If I work with it for a while, one day, so it will work. But as I told you
sometimes my brother might help me. [Participant G; group 2]

Although participants were positive toward using the tool, they believed that they needed time and training to learn and understand its different parts:

I think it [the tool] is very good. I need quite a long time. You need to sit and feel it. What is this so that you understand it?...You first should learn a little bit, then I think it is going to be pretty easy to understand and figure out what you want it to be like...I actually think that I am able to do it [use the tool]. Because now I do not have anything like this here, So It would be really great to be able to know how everything works. [Participant G; group 2]

It takes time, I have noticed, but you have to dig to get the answer. You need to test it; it takes time for everyone. [Participant I; group 2]

The Appropriate Time for Using the Tool After Stroke Occurrence

The patient journey model developed in our previous study [20] consists of different phases, such as “At rehab clinic” and “At home,” and events, such as “Discharge from hospital,” “Discharge from rehab clinic,” “Coming home,” and “Clinical encounters.” A phase may include several events in the patient journey. The participants in this study were in the “At home” phase. Some participants (2/9, 22%) believed that they would have benefited from the tool if they had access to it earlier in their care and rehabilitation journey as they had more severe communication difficulties in the beginning:

It would have been better [to have access to the tool] a bit earlier. I think when I was much worse when I could not talk at all, so it had been really great to look at the curve [the goal attainment scale curve]...In the beginning a lot was happening, three to four months, and then I noticed with my friends I improved a lot so I would benefit more [from the tool] then. I told my speech therapist that I want to have everything, so I am the kind of person that would love to have it [the tool]. [Participant A; group 1]

Design Implications for Patients With Brain Fatigue and Concentration Difficulties

Overview

Many patients experience brain fatigue and concentration difficulties after a stroke. Involving this group of patients in the design process and evaluating the care and rehabilitation plan together with patients who also had aphasia was challenging and required skills and knowledge about both design principles and physical and cognitive disabilities. On the basis of the data collected from the usability tests of the tool, different design implications were identified and are presented in the following sections.

Information Overload

As many stroke survivors with aphasia experience difficulties reading and understanding large amounts of text, designing any screen-based application such as our care and rehabilitation plan requires care. A home page with different icons contained administrative and health-related information. Some of the participants (3/9, 33%) were satisfied with the quantity of information on this page, but most of the participants (5/9, 55%) would like to have the possibility to choose only the information amount that met their needs:

It is easier to choose this [home page], but it is difficult to do other things. It [the home page] was great, it is very easy to use, e.g., rehabilitation, calendar and information about me, it is very good to find [this kind of information]. [Participant C; group 1]

It was pretty hard the whole thing in the beginning. It was a lot in this page [home page]...it is a little too much information (in home page). It would certainly be enough with five [icons]. It is too much. [Participant F; group 1]

It is good, you get help from these different things [information]. [Participant E; group 1]

Although the care and rehabilitation plan was designed based on the information needs of postdischarge stroke survivors, some participants in this study (2/9, 22%) did not find all information amounts in the home page necessary:

Indeed, reminder is necessary. Assistive tools, it is good, but I did not know what it is...You do not know what assistive tools means. I do not have any assistive tools so I do not need it, so I would not click it here. [Participant A; group 1]

The home page includes health and administrative information (Figure 3). Participants would like to be able to choose the necessary information based on their needs. They wished that the home page would include less information.

As some stroke survivors have bad eyesight, it was difficult to follow everything on different pages, for example, the overview page of the tool. However, big icons on the home page were perceived as easy to follow and access:

I actually think that it is good. Large buttons, so it is good. Even though I did not find one thing but like I said it was the first time [that I used the tool]. But it is good when they are so big [the buttons]. For some of us see a bit bad. We might see a bit like this on one side and this on the other side. So perhaps smaller icons...but this is great indeed. [Participant G; group 2]

The amount of information on the overview page was also something that participants explained as difficult to understand and follow. The quantity of information on the overview page increases as the number of problems, goals, and activities increases (Figure 4). Some participants (3/9, 33%) were not satisfied with the increased amount of information on different pages, particularly on the overview page:

It is a bit like this: eee...wait, what is it now...? You wonder what all this is about and what is this...and I might have 6-7 goals and it gets very blurred. [Participant G; group 2]
Complexity of Concepts

Almost none of the participants understood the difference between simple and SMART goals in the rehabilitation plan (Figure 5). Participants had different ideas about what a simple or SMART goal could mean. The difference in the description of the goals shows that the concepts used in the care and rehabilitation plan were not easy and clear to the participants.
The participants wished to have some clarification on different terms and concepts on every page of the tool:

You can do simple goals e.g., at home, but I do not know what SMART goals mean. [Participant C; group 1]

Results/goal achievement sounds really good. Mood history does not sound good. [Participant F; group 1]

I do not know what a simple goal and a smart goal are...show simple goal?! What is this? Show smart goal?! You get so confused, and then my goals?! [Participant A; group 1]

For smart goals you have a deadline. [Participant E; group 1]

Simple and smart goals should be clarified, what is what...I do not understand [ICF codes]. [Participant F; group 1]

Simple goals you do without thinking about it, e.g., opening a door. You do it automatically. You do simple goals every day and smart goals are in future. It is about time. [Participant I; group 1]

**Figure 5.** An overview of simple and Specific, Measurable, Achievable, Relevant, and Time-bound (SMART) goals (English translations of some necessary parts of the figure have been added). GAS: goal attainment scaling.

**Visualization**

Different parts, for example, goals and mood history of the electronic care and rehabilitation planning tool for postdischarge stroke survivors, were visualized using graphs, symbols, images, color-coding, and icons. The patients’ opinions on the visualization are presented in the following sections.

**Using Graphics**

Most participants (6/9, 66%) appreciated the graphs, symbols, and icons used in the care and rehabilitation planning tool. They believed that graphs and symbols supported them in understanding the process better:

I think it [mood history] is perfectly fine with such symbols, because you do not need to write a lot. [Participant A; group 1]

The image (GAS curve) is good, you understand what happened immediately. [Participant F; group 1]

Very good, big icons, different colors, text and images. [Participant I; group 2]

Some participants (3/9, 33%) believed that the tool could include larger letters and figures. They had difficulties finding the necessary information for accomplishing some of the tasks:

[Headings] should be bigger e.g., activities, goals, problems. Large letters, so you can see them right away. I searched and did not see them, and I try the whole page and so on. [Participant A; group 1]

I did not see this [the assessment point of the smart goal] perhaps you can see it if it was bigger. Particularly if it is bigger, now I cannot see it at all. I looked a lot on words [the menu at the top of the page], I searched a lot. [Participant I; group 2]

**Results Visualization**

Despite the fact that showing positive progress in goal achievement curves and mood history can motivate the patients to continue with their activities, negative results can also be a motivation factor that provides patients with an overview of their weaknesses. Goal achievements can be visualized in the tool as defining SMART goals and assessing them makes it possible for the user to obtain an overview in the form of a GAS curve (Figure 6):

Actually, it is good [negative results in GAS curve], for me it would be good to know what happens. [Participant F; group 1]

It is great for sure, to understand how you have felt for a long time. It is better [to have it] so you can do something to make it a little better. [Participant B; group 1]
It [mood history] is usually up there and if it looks like this [down curve] then I think what I have done, so I would think. [Participant A; group 1]

Well, you have to dig, and then you seek after the truth when you know it. You know about your weaknesses, so this is not something new, you get it confirmed what you already know. You do something about the problem. [Participant I; group 2]

Most participants (5/9, 55%) appreciated the mood history in the tool and were positive toward using the mood history in the care and rehabilitation planning tool. They believed that using this feature was easier than using other parts of the tool and liked the idea of having an overview of their mood history over time:

It is very good; it is much easier than the other parts and you understand how it works. [Participant C; group 1]

However, a participant believed that he would benefit more from a fatigue history than from a mood history (Figure 7):

For me it is fatigue. It does not happen so much with my mood, but it happens with fatigue. [Participant E; group 1]

Figure 6. An example of a goal attainment scaling (GAS) curve (English translations of some necessary parts of the figure have been added).

Figure 7. An overview of the mood history in the care and rehabilitation planning tool (English translations of some necessary parts of the figure have been added).

Using Color-Coding

In the tool, the goals, activities, and problems are linked in My rehabilitation using color-coding. Almost none of the participants understood the color-coding part of the tool before the description by the moderator. After clarification by the moderator, the participants were quite satisfied with the
color-coding in the tool and believed that it helped them distinguish between different care professionals (Figure 8):

It is great, e.g., if I have occupational therapist then I have it in yellow. [Participant C; group 1]
I think the color-coding is really good. [Participant F; group 1]

Preferably bigger icon for rehabilitation, very large. It is quite small, there is a lot of space here, but it is good that it is same color and goals. Color-coding facilitates and helps a lot. Preferably bright colors, you can use different shades. [Participant I; group 2]

**Figure 8.** An overview of the connection between different problems, goals, and activities using color-coding.

---

**Perceived Usefulness**

Participants believed that having access to an electronic care and rehabilitation planning tool providing them with an overview of, for example, their problems, goals and activities, and administrative and health-related information would be useful for them throughout their care and rehabilitation processes:

I think it is good, because you can do it on computer, or it can be on TV and phone and because you know what you can do as well. Because I have different things, like for physiotherapy I go to Farsta (an area in south Stockholm) and then I have my speech therapist in Liljeholmen (an area near to central Stockholm) and with another care provider. Here, everything is gathered, and I think it is good, and then I have Alma courses and then you have everything here [in the tool]...Everything is gathered in one place. [Participant G; group 2]

Participants believed that having access to the necessary information in the right place would motivate and support them in their planning throughout the care and rehabilitation processes:

It is good, it makes the life easier, a lot, and then above all, a very good overview, everything is there [information that patients need], e.g., my care contacts...and then my rights and responsibilities, it was interesting for me. You get information in the right place. It is fast. It is fast and easy in one place. [Participant I; group 2]
[It is] available, motivational, at the same place, fun. [Participant A; group 1]
It is nice [to have access to the tool], and it is all I need [the administrative and care related information]. [Participant B; group 1]
To be able to plan [having overview of goals and activities] and I can find everything I need [administrative and care related information]. [Participant C; group 1]

As many stroke survivors have memory loss, some participants (2/9, 22%) believed that using the tool would be helpful for remembering what activities they had done:

Yes, it can be good, and it might be better in retrospect when you can go back and see what you have done before e.g., that you have exercise walking for a while and for me it is to read etc. so you remember what you have done. [Participant E; group 1]
Discussion

Principal Findings

Despite the fact that the participants were positive toward using the care and rehabilitation planning tool, the results showed that the tool was not effective for stroke survivors with aphasia. Participants in group 1 had challenges in using the tool and did not accomplish the tasks successfully. However, those in group 2 managed to accomplish most of the tasks receiving minor support from the moderator.

Participants mentioned that they needed time and training to understand the different parts of the tool. They even mentioned that they would have benefited from the tool if they had had access to it earlier as they had severe communication difficulties at the beginning of their care and rehabilitation processes.

As participants in this study had aphasia and difficulties reading and understanding large amounts of text, they wished to have less information on the home page and the possibility to choose only the information that met their needs. They liked the big icons on the home page but did not appreciate the increased amount of information on different pages of the tool. The use of concepts such as simple and SMART goals in the tool was not perceived positively, and the difference between them was not clear to the participants. Regarding visualization, most participants (69, 66%) appreciated the graphs, symbols, and icons used in the tool as they believed that they helped them understand the process better. The participants even believed that visualization of their mood history along with positive or negative progress in the goal achievement curves could be motivating factors for them to continue their activities and strengthen their weaknesses. However, color-coding used for connecting the goals, activities, and problems in My rehabilitation in the tool was not easily understood by the participants. Regarding the usefulness of the tool, the participants believed that having access to the tool would be useful to them throughout their care and rehabilitation processes as it provides necessary information that would motivate and support them during the entire journey.

Comparison With Prior Work

Previous studies on the use of ICT for supporting patients with aphasia have focused on aphasia therapy, for example, helping patients with their fluency and voice disorders and identifying and determining management strategies for dysphagia [51,52]. However, the care and rehabilitation planning tool in this study focused on supporting this group of patients by providing them with an overview of their rehabilitation plan along with giving them access to necessary information that meets their needs throughout their care and rehabilitation journey. To our knowledge, there is no study focusing on the care and rehabilitation planning process for patients with aphasia and no study involving this group of patients in the evaluation process of an ICT tool. Martin et al [53] evaluated a software tool designed for older adults with aphasia and identified, for example, different design problems related to usability from speech and language therapists’ perspectives. In another study, Reeves et al [54] evaluated a multimedia application for patients with aphasia with 20 speech and language therapists. However, in our study, we evaluated the care and rehabilitation planning tool with patients with aphasia themselves and identified usability problems and the usefulness of the tool from the patients’ perspective.

In a previous study, care professionals’ perceived usefulness of the tool was examined using the Unified Theory of Acceptance and Use of Technology [47]. The results showed that, except for challenges such as time limitation and responsibility issues of the tool, care professionals were positive toward the tool and its potential usefulness throughout the care and rehabilitation processes of postdischarge stroke survivors. The results of this study also showed that despite some challenges, such as usability problems with the tool and patients’ several cognitive and physical disabilities, the patients were positive toward using the tool and appreciated its usefulness in tracking, for example, goals and activities, along with accessing necessary information. The researchers in this study were also aware of issues related to, for example, the responsibility of the system and technical support that need to be considered before the implementation of the system. However, the results of this study regarding participants’ satisfaction and their need to access necessary information and have an overview of their rehabilitation plan showed that designing and evaluating appropriate eHealth services is an increasing necessity in the care processes of patients with chronic illnesses.

Despite the fact that the tool was designed using a user-centered design approach with stroke survivors, there are still many design implications that need to be considered when designing for and with patients with chronic illnesses, particularly stroke survivors who have several physical and cognitive disabilities. The results of this study confirmed the general design implications, for example, information overload and visualization, for patients with cognitive disabilities discussed in previous studies [35,55]. The results showed that the tool designed for stroke survivors might not be effective for patients with aphasia after stroke. Therefore, involving this group of patients in the design process is of great importance even though it might be challenging for the research group and also for the patients. It is also of great importance to involve this group of patients in the evaluation process. Although involving this group of patients in this study was challenging and required skills regarding moderating the usability tests because of the patients’ speech and communication problems, it resulted in a better understanding of how to design appropriate eHealth services for this patient group.

One of the challenges in designing eHealth services for patients with disabilities, in this case, stroke survivors, is to consider their mental fatigue, neglect, and concentration difficulties. The amount of information on different parts of the tool was a problem during the tests. Most participants (69, 66%) became exhausted and frustrated using the tool. Some of them were distracted by the available information and started reading rather than performing the tasks in the test. Despite all these challenges and difficulties for participants and the research group during the tests, the participants were interested in using the tool as they believed that having access to necessary information gathered in one place and keeping track of their rehabilitation...
process are of great importance throughout their care and rehabilitation journey.

Although the results of this study showed that the tool is not effective for stroke survivors with aphasia as only 17% (1/6) of the participants in group 1, who in addition had a computer science background, was able to accomplish most of the tasks, we cannot draw any conclusions about the effectiveness of the tool based on this participant’s performance as their background might have affected the way the tasks were accomplished. However, providing some minor support resulted in greater effectiveness of the tool as most of the tasks were accomplished successfully by the participants in group 2. Overall, it was easier for most participants in both groups to perform the tasks related to finding information, but tasks related to the establishment of a rehabilitation plan consisting of several steps required more effort from the participants. However, for most patients, it was difficult to add their own goals and activities. Having support from participants’ next of kin was mentioned during the tests. Therefore, it might be easier for these patients to establish a rehabilitation plan together with their next of kin or health care professionals.

Limitations
As only a few participants (3/9, 33%) were able to accomplish most of the tasks successfully, the care and rehabilitation planning tool should be more intuitive and adaptive, particularly when establishing the rehabilitation plan. Some participants in this study (3/9, 33%) mentioned that they would learn how to work with the tool if they had enough time and the opportunity to use it frequently. Therefore, it is of great importance to also evaluate the tool in the future to study its effectiveness after a certain period.

As this study was designed and conducted in 2016, the research group did not have access to the literature focusing on developing aphasia-friendly technology that has been published in recent years. To improve the system, in the next design or update round of the studies by Davoody et al [41], these studies [56-59] will be considered. These studies can be used to incorporate different elements of this research into the development or adaptation of the care and rehabilitation planning tool to assess its usability with people with aphasia.

Conclusions
Although the care and rehabilitation planning tool is not effective for postdischarge stroke survivors with aphasia, it can be usable for this group of patients provided they receive some minor support from their next of kin or care professionals in a neurology team. However, eHealth services and HISs designed for these patients should be more adaptable and flexible to provide patients with appropriate functionalities and features and be useful and easy to use.

Despite the fact that involving patients with chronic illnesses with several physical and cognitive disabilities, particularly patients with aphasia, in the evaluation processes is a challenge because of their communication difficulties, their input for developing appropriate HISs and eHealth services is crucial. As the challenges existed throughout the user-centered design approach used for designing the care and rehabilitation planning tool, the design and evaluation processes should be adapted, considering these challenges.

Evaluating the care and rehabilitation planning tool from the perspective of this patient group provides insights into some of their information and communication needs. The results of this study show that patients with aphasia could benefit more from the tool if it could be adapted to their needs. However, further research is needed to confirm that the adjusted tool could be useful for this patient group.

In addition, to give different patient groups the opportunity to adopt the care and rehabilitation planning tool for their disabilities, different user profiles can be developed within the tool.

Acknowledgments
The authors would like to thank all the participants in this study. They provided valuable information. The project was partly supported by Vinnova, the Swedish Agency for Innovation Systems (grant 2016-00646), and by Forte, the Swedish Research Council for Health, Working Life, and Welfare (grant 2016-07324). The financier had no influence on the design and content of the study.

Authors’ Contributions
All the authors participated in the study design. ND collected the data. All authors participated in the data analysis and writing of the manuscript.

Conflicts of Interest
None declared.

References


Abbreviations

GAS: goal attainment scaling
HIS: health information system
ICT: information and communications technology
SMART: Specific, Measurable, Achievable, Relevant, and Time-bound
in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Trust and Uncertainty in the Implementation of a Pilot Remote Blood Pressure Monitoring Program in Primary Care: Qualitative Study of Patient and Health Care Professional Views

Evelyn Chew¹, BA, LicArts, MA, PhD; Sok Huang Teo¹, BSc, MPH; Wern Ee Tang¹, MMed; David Wei Liang Ng¹, MMed; Gerald Choon Huat Koh²,³, MMed, PhD; Valerie Hui Ying Teo¹, MMed

¹National Healthcare Group Polyclinics, Singapore, Singapore
²Ministry of Health Office for Healthcare Transformation, Singapore, Singapore
³Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore

Abstract

Background: Trust is of fundamental importance to the adoption of technologies in health care. The increasing use of telemedicine worldwide makes it important to consider user views and experiences. In particular, we ask how the mediation of a technological platform alters the trust relationship between patient and health care provider.

Objective: To date, few qualitative studies have focused on trust in the use of remote health care technologies. This study examined the perspectives of patients and clinical staff who participated in a remote blood pressure monitoring program, focusing on their experiences of trust and uncertainty in the use of technology and how this telehealth intervention may have affected the patient-provider relationship.

Methods: A secondary qualitative analysis using inductive thematic analysis was conducted on interview data from 13 patients and 8 staff members who participated in a remote blood pressure monitoring program to elicit themes related to trust.

Results: In total, 4 themes were elicited that showed increased trust (patients felt reassured, patients trusted the telehealth program, staff felt that the data were trustworthy, and a better patient-provider partnership based on the mutually trusted data), and 4 themes were elicited that reflected decreased trust (patients’ distrust of technology, clinicians’ concerns about the limitations of technologically mediated interactions, experiences of uncertainty, and institutional risk).

Conclusions: Managing trust relationships plays an important role in the successful implementation of telemedicine. Ensuring that trust building is incorporated in the design of telehealth interventions can contribute to improved effectiveness and quality of care.

Introduction

Background

An aging population worldwide and the consequent increased workload for health care systems have led to growing interest in technological innovations that can potentially lessen the strain on overtaxed health care systems [1,2]. Remote monitoring of blood pressure (BP) in patients with hypertension is an area that shows promise [3,4]. Studies to date featuring remote BP monitoring have reported positive outcomes in terms of
acceptability and improved ability to manage one’s health, including in a predominantly minority, lower-income older adult population [5-7]. Home BP monitoring is now recommended as part of treatment in the clinical guidelines of several countries [8].

However, despite interest and enthusiasm on the part of health care providers, the mainstreaming and long-term sustainable implementation of such telehealth services is often fraught with challenges [9-13]. This is at least in part because the successful adoption and implementation of any telemedicine intervention depends heavily on human factors such as trust and acceptance, the lack of which can impede or even derail a program [9,10].

Trust is crucial to the health care provider-patient relationship and is very much at stake in the implementation of digital health. To begin with, the patient-provider relationship is fundamentally based on both interpersonal and institutional trust [14-16]. A trusting relationship with one’s health care provider is linked to better adherence to treatment and perceived effectiveness of care, whereas lack of trust is associated with lower rates of care seeking and appropriate treatment [15,17]. Telemedicine “necessarily alters the context of the traditional face-to-face physician-patient trust-based relationship,” in a shift that “may transform the substance of that relationship” [15]. Social shaping of technology theories tell us that technology design shapes user behavior; users, in turn, both shape and are shaped by the technology as they interact with it and within the larger system [18,19]. With regard to telemedicine, many questions arise: how does the patient-provider relationship change on an individual level because of the mediation of technology? How might the patient’s trust in the health care institution be affected? How much do patients and health care professionals trust the technology itself?

Presti et al [20] define trust as “an evolving, contextual and composite belief that one principal (trustor) has that another principal (trustee) will perform certain actions with certain expected results, when not all information about those actions is available.” More specific to e-services, the definition by Grandison [21] narrows this down to “the quantified belief by a trustor with respect to the competence, honesty, security and dependability of a trustee within a specified context.” Nevertheless, trust is a difficult notion to conceptualize and operationalize, and a vast array of conceptual categorizations and models of trust appears in the literature on trust and digital health, spanning psychology, management studies, IT studies, and health care research [21-30].

The early interdisciplinary model of trust by McKnight and Chervany [31] distinguished between dispositional, institutional, and interpersonal trust. Dispositional trust is intrapersonal, something that lies within a person; institutional trust is impersonal, grounded in situations or structure; whereas interpersonal trust refers to “trust in specific others.” In this early model, trust in technology is grouped under institutional trust.

The question of trust quickly rises to the fore in any technology-mediated service provider relationship. The technology acceptance model [32] initially focused on perceived usefulness and perceived ease of use but was soon expanded by researchers to include personal dispositions to trust, institution-based trust, and previous internet experiences, as well as user beliefs, attitudes, and intentions regarding the web-based environment [24,33,34]. Similarly, the Unified Theory of Acceptance and Use of Technology by Venkatesh et al [35] quickly expanded to include dimensions of trust. Elaborating on this model, Pal et al [36] found that, in the context of health care, perceived trust, technology anxiety, and expert advice were important factors for older adults’ acceptance of the Internet of Things and smart home technology. Deng et al [37], testing an extended version of the technology acceptance model, incorporated the role of trust and found that trust was the most important factor in patients’ adoption intention, whereas Arfi et al [29] found that perceived risk mediated perceived trust.

Finally, the eHealth Trust Model by Shen et al [25], which directly focuses on eHealth, integrates the Antecedent, Privacy Concern, and Outcome model and the Web-Trust Model [38]. In total, 6 antecedents to trust are listed: privacy experience, eHealth awareness, health care perception, demographic, technological savviness, and culture [25]. Clearly, trust is an important component of any system in which health, humans, and technology interact.

**Objectives**

The intervention under study involved patients with chronic hypertension using a remote BP monitoring system to measure and upload their BP readings to a secure remote server, monitored periodically by the health care team. Follow-ups occurred through telephone consultations. This study sought to shed light on some ways in which such a program might positively or negatively affect trust relationships in health care. Drawing from existing literature and observing the human and nonhuman actors involved in the telehealth program led us to deduce that 3 kinds of trust were of relevance: interpersonal trust, institutional trust, and human-technology trust [16,39,40]. Interpersonal trust refers to the trust between a patient and the individual health care professional. This trust is not only one-way, from patient to health care professional; in a home-based telehealth intervention, the professional must also trust that the patient will play their part. Moreover, interpersonal trust between health care professionals is involved when health care professionals must work together as a team in implementing the telehealth intervention. Institutional trust is the trust that patients place in the health care institution as a whole [41,42]. Human-technology trust [30] relates to patients’ and staff’s individual attitudes of trust toward the telehealth technology. The trust relationships implicated among the actors in this intervention are illustrated in Figure 1.

To date, few qualitative field studies have focused on the issue of trust in telehealth. Most existing studies on trust have carried out general surveys or built frameworks based on conceptual analyses. Thus, the value of this study was to use data available from our field quasi-experiment to extend existing findings about how trust in the patient-health care provider relationship is affected when a telehealth intervention is introduced.
Epistemologically, this study took a broadly critical realist and social interactionist approach [43,44]. Critical realism links the examination of structure and agency (germane to critical theory) with observable realities, thus remaining close to ground-level data. It also acknowledges that reality is an open system made complex by multiple, potentially nonreplicable causal mechanisms [45]. Social interactionism highlights that social realities are created and given meaning through human beings’ interactions with one another. Studying the question of trust in the patient-provider telehealth relationship through these lenses allowed us to interrogate social meanings and interactions and thereby elucidate the implications of such a program on trust in patient-provider relationships.

Figure 1. Trust relationships involved in the remote blood pressure (BP) monitoring telemedicine intervention.

Methods

Overview

An interventional, quasi-experimental remote BP monitoring program was conducted in a polyclinic in Singapore from September 2018 to September 2019 involving 217 patients. Patients with hypertension were assigned to either a control arm or an intervention arm.

Patients in the intervention group were given a Bluetooth-enabled home BP monitor (TaiDoc Technology FORA P20b Blood Pressure Monitoring System) and a mobile data network–connecting gateway device (Phicomm Clue C230) that connected to a secure remote server. During a one-on-one in-person training session, patients were instructed on how to use the cuff and equipment to properly measure and upload their BP readings to the server. They were tasked to do this at least once a week from home over a period of 6 months to a year. For the duration of the study, care managers who were nurses periodically reviewed the patients’ BP readings. Instead of in-person visits, patients whose BP was well controlled reviewed their condition through telephone consultations with their care managers (scheduled teleconsultations). If unexpectedly high readings were detected, care managers would contact patients to check on their well-being (unscheduled teleconsultations). Where clinically indicated, medications were adjusted over the phone after consultation with a physician. Quantitative and qualitative data were gathered and have been reported elsewhere (Teo, S, unpublished data, October 2022) [46]. This study is a secondary analysis of the qualitative interview data from patients in the intervention group.

Setting

Singapore is a small, highly urbanized country in Asia with >5.7 million inhabitants. The country has an internet penetration rate of >81% [47], making it ideal for telemedicine, which is becoming increasingly popular [48]. The primary health care scene in Singapore comprises public and private institutions. Public health care is subsidized, with physicians in polyclinics—which are the public primary health care institutions—taking on a large share of the treatment of chronic illnesses [49]. Our study was set in a polyclinic in central Singapore.

Participants

Participants were patients with hypertension from the intervention arm of the study [46] and staff who were involved in the program. Patient interviewees were referred by attending clinicians; staff involved in the research study were invited to participate by members of the research team. Patient participants had been in the remote BP monitoring program for at least 6 months. Of the 20 patients and 10 staff approached, 13 (65%) patients (n=8, 62% male and n=5, 38% female, with ages ranging from 35 to 73 years) and 8 (80%) staff members (n=2, 25% physicians; n=3, 38% care managers; n=1, 12% senior
nurse clinicians; and n=2, 25% care coordinators) agreed to be interviewed. Participant demographics are listed in Tables 1 and 2.

Table 1. Patient participant demographics.

<table>
<thead>
<tr>
<th>Participant ID</th>
<th>Sex</th>
<th>Age (years)</th>
<th>Education level</th>
<th>Occupation</th>
</tr>
</thead>
<tbody>
<tr>
<td>F007</td>
<td>Male</td>
<td>47</td>
<td>Tertiary</td>
<td>Nursing home manager</td>
</tr>
<tr>
<td>F009</td>
<td>Female</td>
<td>47</td>
<td>Secondary</td>
<td>Cashier</td>
</tr>
<tr>
<td>F029</td>
<td>Male</td>
<td>50</td>
<td>Tertiary</td>
<td>IT manager</td>
</tr>
<tr>
<td>F021</td>
<td>Male</td>
<td>58</td>
<td>Tertiary</td>
<td>Teacher</td>
</tr>
<tr>
<td>F026</td>
<td>Male</td>
<td>49</td>
<td>Tertiary</td>
<td>Engineer</td>
</tr>
<tr>
<td>F035</td>
<td>Male</td>
<td>67</td>
<td>Secondary</td>
<td>Part-time consultant</td>
</tr>
<tr>
<td>F022</td>
<td>Male</td>
<td>64</td>
<td>Tertiary</td>
<td>Senior management</td>
</tr>
<tr>
<td>F096</td>
<td>Female</td>
<td>46</td>
<td>Secondary</td>
<td>Data entry clerk</td>
</tr>
<tr>
<td>F099</td>
<td>Female</td>
<td>64</td>
<td>Preuniversity</td>
<td>Not working</td>
</tr>
<tr>
<td>F118</td>
<td>Female</td>
<td>67</td>
<td>Primary</td>
<td>Part-time office cleaner</td>
</tr>
<tr>
<td>F119</td>
<td>Male</td>
<td>58</td>
<td>Preuniversity</td>
<td>Warehouse manager</td>
</tr>
<tr>
<td>F110</td>
<td>Male</td>
<td>35</td>
<td>Tertiary</td>
<td>Teacher</td>
</tr>
<tr>
<td>F122</td>
<td>Female</td>
<td>73</td>
<td>Tertiary</td>
<td>Not working</td>
</tr>
</tbody>
</table>

Table 2. Staff participant demographics.

<table>
<thead>
<tr>
<th>Participant ID</th>
<th>Sex</th>
<th>Job title</th>
<th>Role in telemedicine program</th>
</tr>
</thead>
<tbody>
<tr>
<td>S001</td>
<td>Female</td>
<td>Care coordinator</td>
<td>Train participants to use remote BP&lt;sup&gt;a&lt;/sup&gt; monitor; provide follow-up technical support</td>
</tr>
<tr>
<td>S002</td>
<td>Female</td>
<td>Care manager</td>
<td>Teleconsultation; monitor BP readings</td>
</tr>
<tr>
<td>S003</td>
<td>Female</td>
<td>Care manager</td>
<td>Teleconsultation; monitor BP readings</td>
</tr>
<tr>
<td>S004</td>
<td>Female</td>
<td>Family physician</td>
<td>Approve medication adjustments; counsel patient on medication changes</td>
</tr>
<tr>
<td>S005</td>
<td>Female</td>
<td>Care manager</td>
<td>Teleconsultation; monitor BP readings</td>
</tr>
<tr>
<td>S006</td>
<td>Male</td>
<td>Family physician</td>
<td>Approve medication adjustments; counsel patient on medication changes</td>
</tr>
<tr>
<td>S007</td>
<td>Female</td>
<td>Care coordinator</td>
<td>Train participants to use remote BP monitor; provide follow-up technical support</td>
</tr>
<tr>
<td>S008</td>
<td>Female</td>
<td>Senior nurse clinician</td>
<td>Support back-end coordination and implementation</td>
</tr>
</tbody>
</table>

<sup>a</sup>BP: blood pressure.

Procedure

Interviews were carried out by study team members (female research fellows ECAL and TSH, who were experienced in qualitative research) face-to-face in a quiet room in the polyclinic before or after patients’ appointments. Patient participants were asked about their experiences of living with high BP; their thoughts on the telehealth program; and their experiences with the remote BP monitoring equipment, teleconsultations, and remote medication review (see Multimedia Appendix 1 for the interview topic guide). Staff participants were asked about their experiences with the telehealth program, focusing on their specific role (onboarding, teleconsultation, and medication adjustment) in carrying out the program. Patient and staff interviews lasted slightly less than an hour each. The interviews were audio recorded and transcribed verbatim. Interviewers had no direct or personal working relationship with interviewees apart from this study.

Data Analysis

Using a Microsoft Excel (Microsoft Corp) spreadsheet and the transcribed interviews, an initial round of inductive thematic analysis was conducted by ECAL and TSH to elicit main themes from the interview data (Teo, S, unpublished data, October 2022). In the process, the question of trust arose as a theme that merited more detailed study. A secondary thematic analysis was run on the transcripts, focusing on the question of trust to further draw out other aspects of the theme. Secondary analysis of qualitative data is appropriate for cases where a researcher wishes to broaden and deepen in knowledge using data that have already been gathered [50-53]. As this is a secondary analysis, the principle of data sufficiency rather than data saturation was applied [54,55].

Ethics Approval

The study was approved by the relevant institutional ethics board (Domain-Specific Review Board 2018/00785).
Results

Overview
We found 4 themes reflecting increased trust among the parties involved in the telehealth program and 4 themes that reflected reduced trust. Themes reflecting increased trust were as follows: patients felt reassured, patients trusted technology and the telehealth program, clinicians trusted the data generated, and a sense of partnership arose from the mutually trusted data. Themes that displayed reduced trust were as follows: patients’ distrust of technology, clinicians’ concerns about the limitations of technology, experiences of uncertainty, and institutional risk. Although it is not possible to list quotations from all participants, Textbox 1 lays out the themes along with the participants whose views support them.

Textbox 1. Themes elicited.

- Themes reflecting increased trust
  - Patients feel reassured that “someone is monitoring”: participants F007, F009, F035, F096, F110, F122, S002, S003, S005, and S006 (clinicians’ perceptions that patients felt reassured)
  - Patients’ trust in technology and telehealth: participants F007, F021, F022, F029, and F035
  - Clinicians’ trust in technologically generated data: participants S004 and S006
  - Mutually trusted data support patient-clinician partnership: participants S004, S005, S006, S008, F007, F009, F026, F110, and F122

- Themes reflecting decreased trust
  - Patients’ distrust of or discomfort with technology: participants F021, F099, F118, F119, S002, S006, S007, and S008
  - Clinicians’ concerns regarding the limitations of technology-mediated interactions: participants S004, S005, and S008
  - Experiences of uncertainty: participants F007, F009, F022, F029, F119, F122, S002, S005, and S008
  - Institutional risk: participants S007 and S008

Patients Feeling Reassured That “Someone Is Monitoring” (Interpersonal and Institutional Trust)
A dominant theme that emerged was that patients felt reassured that they were being closely followed up with by their health care team. This gave them a sense of security and increased their trust in both the health care professionals and the health care provider as an institution. Patients felt that they could “relax,” “knowing that at the other end, there is somebody looking at your [BP] readings” (participant F122):

> I know that the [public healthcare institution] has my records and maybe if there’s any irregular kind of BP, hikes or something like that, they all might call me. [participant F007]

> I feel great, ya. At least I know that the polyclinic is keeping track of my blood pressure and then they’ll always make an effort to call us. [participant F110]

A participant reported feeling “happy” when he was called by the clinic after submitting an unexpectedly high BP reading. He recognized that he was not alone in being concerned about his BP and felt supported in managing it:

> ...my blood pressure went suddenly went [up to] 150! They called me up...[laughs] they call me up, I feel happy!!...If I do like [before the program], take [my BP reading as] 150, and just leave it behind, don’t care, 150! Then if [if it were to] drag longer, then the blood pressure keeps on going up, if there’s a problem we don’t know also! [participant F035]

Participants felt that they received more tailored guidance because of the intervention. Participant F035 added that, if not for the program, he would likely have ignored the high BP reading, lacking information on how to proceed, whereas the call from the clinic both reassured him and gave him specific steps to follow.

Another participant received a call the day after she submitted a high BP reading. She did not pick up the call as she was busy, leading the clinician to call several times. She was impressed by the swiftness and effort the staff made to contact her about her abnormally high BP, and this prompted her to take her condition more seriously:

> When they called, then I know it’s actually serious for them. [participant F096]

Others reported that the individualized guidance during the program on how and when to measure their BP correctly and avoid false measurements improved their ability to self-manage their condition.

The staff interviewees echoed this view:

> The patients, most of them actually seem quite appreciative of it. Like they think it helps them—someone is monitoring, maybe it gives them reassurance. That we’re looking into readings. [participant S003, care manager]

> Compared to usual care, because [in] usual care you don’t really keep monitoring their blood pressure...whereas this [program], as and when you see slightly borderline high, you will just call. Ah, so they know that you are there. [participant S005, care manager]
Patients felt supported in their health care management because of the perceived closer follow-up on the health care provider’s part. Interestingly, this trust was not linked to specific health care professionals but was often identified with the polyclinic staff as a whole using the generic pronoun “you all.”

**Patients’ Trust in Technology and Telehealth (Human-Technology Trust)**

Among interviewees, attitudes toward technology—and, hence, trust in the remote BP monitoring program—varied widely. Some were enthusiastic about the usefulness of the program and appreciated the feedback from the BP monitor and the calls from the clinic, which helped them be more consistently aware of their health condition and its management; others were more apprehensive. The interviewees’ occupations also influenced what they thought of telemedicine. In total, 10% (2/21) of the interviewees—a manager of a nursing home and an IT professional—were particularly supportive of telemedicine and took a systemic view, arguing that technology not only could but ought to be leveraged to create efficiencies for the health care system:

> A lot of things can be done by yourself, rather than needing a face-to-face with doctors. Sometimes you need [it], but not all the time. Sometimes online is good enough...If let’s say I have a particular question, if I can text or whatever some questions, [and] somebody can reply, [that] can be good as well. [participant F022]

**Clinicians’ Trust in Technologically Generated Data (Human-Technology Trust)**

In total, 100% (2/2) of the physicians interviewed appeared to trust the readings from the remote BP monitor more than the data generated by the previous system in which patients manually recorded their BP on paper. One physician noted that the remote BP monitoring data helped root out false “white coat” hypertension readings as patients’ readings taken at home would better reflect their BP in ordinary life. The other felt that the remotely generated readings were “more accurate” because of the following:

> ...they can’t alter it. The old [system], you can write down, you can erase it. You can write down a good reading, you can hide the high reading. So when they’re using the manual [record system], sometimes if I ask them further, they actually do have very high readings—they just don’t write it down. [Whereas the remote BP monitoring program takes] multiple measurements. So with this, in a way it’s more accurately reflecting their actual level of control for their blood pressure and...they can’t cheat. [participant S006, physician]

**Mutually Trusted Data Support Patient-Clinician Partnership (Interpersonal Trust)**

As the program required patients to provide sustained readings over a longer period, and as they were themselves involved in measuring and uploading the BP readings, patients themselves tended to trust the readings more than those recorded during their previous clinic visits. This enabled health care professionals to point to an objective and more accurate reading of BP over time, taken in situ in the environmental context of the person’s life.

A physician found that patients in this program were “more receptive” to advice and attributed it to patients’ “extra sense of security” in the patient-health care provider relationship owing to their active participation in the program, which resulted in increased interactions with the health care provider over 6 months:

> Throughout the process of the six months of monitoring, it’s like a two-way thing. They submit the reading, high or low, we help to interpret. And it is not all the time that we push them to increase the medicine; sometimes...we give some compliments, throughout the process...it kind of strengthens the rapport, so they have higher level of trust, I guess...I just feel like it’s easier to talk to them about their [health] management when they come for the subsequent follow-up review. [participant S006, physician]

As patients considered the BP readings they had submitted to be reliable and trustworthy, it was easier for physicians to present the data as evidence to persuade them to alter or begin a medication regimen when necessary:

> ...there are some patients whose [blood] pressure is always a little bit higher, but they always give excuses right? That it’s their stress, they just came [to the clinic] and they were walking...and things like that. So when they go home and realize the [blood] pressure is also high at home, then it’s a little bit easier to convince them that your [blood] pressure is not well-controlled and [there’s a] need to increase the medicine. [participant S004, physician]

Some patients expressed stalwart support for the program, linked to implicit trust in the health care provider and awareness of their own role in actively managing their health:

> I will never stop [participating in this program], because you know why? Right now the clinic is observing your blood pressure, anything [they will] call us, anything. Any problem, we will, you know, get the problem solved by the doctor overseeing. [participant F035]

Actually [when I] signed on to this [program] I was thinking that uh—it’s a way that the polyclinic helps us to monitor blood pressure...in case that one day it really happens that we don’t know that we actually have blood pressure all the way [dangerously high blood pressure]. Because I think a lot of people, they aren’t aware that they do have [high] blood pressure. So this one can keep monitoring, so at least we got the awareness...I was thinking that it was quite a good project. So, have to try to take it up. [participant F009]

In short, the remote BP monitoring program seems to have fostered greater partnership between the health care provider and patient via trust in the telehealth technology. As patients...
trusted the BP data that they had measured and uploaded, they also tended to trust the clinicians’ advice when those data were used as evidence to persuade them to engage in health-sustaining behavior.

However, for trust to grow, the initial rapport needed to have been established previously with a face-to-face consultation:

If I have a patient who has just transferred from a [private] GP and I put him on a machine, he won’t feel comfortable at all, I don’t think the patient will want to do that. [participant S006, physician]

Lack of a previously established trust relationship resulted in longer teleconsultation time as the physician had to spend more time convincing the patient to follow their advice:

[When] rapport is not completely built up...it’s not as simple, even when you do the teleconsult titration [medication adjustment], over the phone we have to talk longer. [participant S006, physician]

Patients’ Distrust of or Discomfort With Technology (Human-Technology Trust)

The success of the aforementioned patient-provider partnership depends in large measure on human trust in technology—patients’ and health care professionals’ trust in the technological system in use. Where this trust is lacking, uncertainty and discomfort result. In total, 4 themes that negatively affected trust relationships (among patients, the health care provider, and the telehealth program) were patients’ discomfort with and distrust of technology, clinicians’ concerns about the limitations of the technology, uncertainties arising from lack of feedback from the program, and concerns about institutional risk.

Not all patients took to the BP monitoring device with enthusiasm; at least 15% (2/13) demonstrated ambivalence toward the program. Despite having agreed to participate in the telemedicine program, a few older patients became very nervous while interacting with the devices. Those who were hesitant about telemedicine found that experiences of failure or perceived failure to accomplish the task of uploading the BP readings correctly exacerbated their uncertainty and apprehension toward the telehealth program. For instance, a female participant aged 67 years was not used to technological devices and had to spend more time convincing the patient to follow their advice:

I started to give myself pressure. When it was time to measure my BP, I would become very nervous, I felt stressed. So my daughter said it’s better to drop out. [participant F118]

For a minority (2/13, 15%) of interviewees, such as participant F118, who had only basic primary education, apprehension was reflected in their reactions to the telehealth program. The same feature valued by some patients—follow-up calls from care providers—caused anxiety for these participants. Participant F099 would also become anxious whenever she recorded a higher BP reading or received a teleconsultation call. She associated calls from clinics and hospitals with bad news and would rather not hear from the health care provider at all:

If somebody call me, means something [is] wrong, I don’t like...So, if they don’t call me, it’s because my reading is good. If my reading is not good, they will call, definitely [to] ask me to increase my medicine. [participant F099]

Moreover, at least one patient (participant F021) felt a need to present a positive result to the health care provider as he felt that the initial higher reading that he obtained was not reflective of his typical BP and he did not want the clinic to call him. To ensure that a good reading was uploaded, he first would measure his BP using his own BP monitor and then repeat the process with the remote BP monitor provided by the polyclinic only when the readings were favorable.

Some patients embraced certain aspects of the program but not others. Although they valued the “extra sense of security” of having their BP monitored remotely (5/13, 38%), some patients (2/13, 15%) disliked the aspect of remote phone consultations replacing physical visits. They lacked trust in the validity of a remote telephone consultation and felt safer seeing a physician face-to-face. A phone consultation was considered a dubious and poor substitute for an in-person consultation. As a result, clinicians reported that some patients would call in to cancel their teleconsultations and show up to the clinic instead for their routine consultations, as they used to do before the program. As one physician (participant S006) pointed out, this was contrary to the purpose of the program, which sought to reduce clinic visits via self-management and remote BP monitoring.

In a similar vein, some patients were reluctant to increase the dosage of their medications over the phone as they lacked confidence in the remote consultations. This was reported by 15% (2/13) of the patient interviewees and 12% (1/8) of the staff interviewees. A staff participant opined that this was because “they don’t see you” (participant S002, care manager). Reflecting a deep-seated uncertainty as to the trustworthiness of remote telephone consultations, an interviewee who rejected the possibility of having his medication adjusted over the phone explained the following:

...If let’s say, they want to increase [my dosage], I would rather come and meet and find out why I need to increase...We are not sure, doctors are busy also. Did they make a mistake or not? This—that is a phobia. Did they make a mistake? You know? Or it may be somebody’s information, but you called the wrong person. So I will—as far as medication is concerned, for my health or any disease I’m suffering, I’d rather have face-to-face...Sometimes, certain things, I don’t feel comfortable talking on the phone. [participant F119]

Finally, a few (3/13, 23%) patients shared their cybersecurity concerns—where their data would be stored and the possibility of leaked personal information. A patient pointed out the
Clinicians’ Concerns Regarding Limitations of Technology-Mediated Interactions (Human-Technology Trust)

Mirroring patients’ uncertainties about whether telemedicine could provide the same level of care as an in-person consultation were clinicians’ concerns about the teleconsultations. Clinicians’ apprehension centered on the inability to ascertain if their messages were correctly received by patients over the phone:

...you must really make sure that they understand... Sometimes when you talk, you think you are quite clear, but the other party’s hearing is a bit [impaired]. And then, they don’t know what you asked them to do. [participant S005, care manager]

Participant S004, a physician, agreed that “sometimes it’s a bit dangerous to do things over the phone” and felt that teleconsultations should be reserved for patients with greater health literacy and adequate social support to avoid miscommunication. Particularly with medication adjustments, a clinician worried about the extended time between in-person visits:

We still want them to come back, we still want to see them [to find out] whether they’re taking [their medications] or not, if there are side effects, do they know when to stop... [participant S005, care manager]

The potential for miscommunication over the phone was also greater than in face-to-face consultations:

We can’t see the body language. Face-to-face, if I know that you are not paying attention to what I say, then I have to repeat, repeat...But if over the phone, I cannot [be sure] that you are actually listening correctly... Then [they] may end up taking [the medication] wrongly. [participant S008, senior nurse clinician]

A care manager noted that it was easier to build interpersonal rapport, elicit information about lifestyle and medication compliance, and clarify doubts with the patient in person. She also highlighted that some patient caregivers were worried about unclear communication over the call and, therefore, would prefer to avoid teleconsultations.

In short, some patients and some staff interviewees had concerns about the limitations of a teleconsultation compared with a face-to-face encounter. Moreover, despite acknowledging the advantages of time savings and convenience, a few interviewees among both staff and patients felt that in-person visits provided more information than telephone consultations.

Experiences of Uncertainty (Human-Technology Trust)

Feelings of uncertainty regarding diverse aspects of the telehealth program marked several interviewees’ responses among both patients and staff. In total, 3 aspects were identified: lack of visibility (clinicians), lack of feedback from the telehealth system (patients), and lack of feedback from health care professionals (patients).

For health care professionals, uncertainty arose from the lack of visibility of certain information owing to the properties of the telemedicine technology. For instance, from the BP readings on the back end of the system, care managers were unable to ascertain the “why” of an abnormal reading—seeing only the BP measurements, they did not know if patients’ high readings were the result of exercise rather than disease. When a participant failed to upload their readings, they were unable to verify whether the readings had indeed been taken but were not transmitted owing to a technical glitch or whether the patient had neglected to do the requisite BP monitoring. Therefore, patients’ irregular submission of data caused concern for clinicians:

Some patients submit readings irregularly, then at the back end, I worry whether the patient is having any problems... Then I start to call them. [participant S002, care manager]

For patients, lack of feedback from the system surfaced as a design flaw in the BP monitoring device that created uncertainty and discomfort. For instance, uncertainty over whether the readings had been uploaded to the server led some patients to send in several readings in a row, leading to multiple recorded readings that mystified the care manager in charge:

...I asked the patient out of curiosity, “How come you measured your blood pressure so many times in a minute? Or in five minutes so many readings?”... They say it seemed like the reading was not transmitted, so they kept re-measuring, re-measuring, re-measuring... so the numbers keep transmitting to us and we get a lot of readings... and I cannot stop them, because it might be true [and] if I stop this practice, I might get no reading here in the end. [participant S005, care manager]

For one participant, uncertainty arose from the lack of a channel to clarify her medical doubts when side effects occurred after having her medication adjusted:

So last week I took the new pill which is a tablet, I kind of feel a bit strange, uh not—not... there’s something that I cannot explain... I would prefer that there is a contact that I can call. Because teleconsult—that through telephone, they may just say, okay you just take and that’s it. But what if I take and I don’t feel quite right?... So my point is, if we are going to go through this program, we won’t come back until maybe six months later or some time, that...
could be a bit too far, especially if I’m on new medication. [participant F029]

In short, both patients and staff experienced uncertainty. For health care staff, this was related to the limited information provided by the telehealth system and the incomplete picture they were able to form of the patient’s state of health. For patients, it was related to limited feedback (from the telehealth system and the uploading process) and to the inability to clarify doubts about their health condition.

Institutional Risk (Institutional Trust)

The health care professionals we interviewed were acutely aware of the risk that a failure in the accuracy of the remote BP equipment might pose to patients’ trust in the health care institution as a whole. A few patients occasionally noted discrepant readings between their own BP monitoring devices and the study equipment, which caused some concern among staff. A staff interviewee was worried that “for anything that turns bad, there might be negative impact, like they lose trust in our treatment because the devices don’t work well or it’s not as accurate as it should—they expect it to be” (participant S007, care coordinator).

Clinicians were also somewhat concerned about overblown patient expectations of what the telehealth program could achieve. For instance, some patients might expect an immediate response from the medical team in the case of unexpectedly high BP readings, which could indicate a medical emergency. Failure to respond quickly in such a case might result in grave medical consequences as well as disappointment and distrust in the health care provider. To prevent such a situation from happening, staff reminded patients of the limits of the program. A staff interviewee stated emphatically that, if 2 subsequent readings were abnormally high, patients should “always give us a call immediately...like, do not wait, do not wait for our call because it is not real time monitoring. And we always emphasize, [it is] not real-time” (participant S007, care coordinator). Although these guidelines were primarily geared toward patient safety, health care staff were also aware of the reputational risk for the institution implied in the telehealth program.

Discussion

Principal Findings

Our analysis showed that patients and staff both felt that the telehealth intervention had an overall positive impact on interpersonal and institutional trust in the patient-health care system relationship. The telehealth intervention was generally well received—patients felt reassured and trusted the technology, clinicians trusted the technology and the patient-generated data, and this enabled greater partnership in patients’ health management. Nevertheless, the intervention also surfaced some underlying anxieties and concerns that patients and staff alike had about the telehealth intervention, viz., some patients’ distrust of or discomfort with technology, clinicians’ concerns regarding the limitations of technology-mediated interactions, patients’ and clinicians’ experiences of uncertainty, and institutional risk. Given the age distribution of patient interviewees as predominantly 40 to 70 years, it should be noted that the findings may reflect the views of this particular demographic, which may differ from the views of younger patients.

The most salient themes in the data related to patients’ trust in individual health care professionals, in the health care institution, and in technology (relationships 1, 4, and 5 in Figure 1) and to clinicians’ trust in technology (relationship 6 in Figure 1). Table 3 summarizes the findings.

Table 3. Summary of themes and trust relationships.

<table>
<thead>
<tr>
<th>Relationship</th>
<th>Type of trust relationship</th>
<th>Trust-facilitating theme</th>
<th>Trust-hindering theme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient-health care professional (1 and 2)</td>
<td>Interpersonal trust</td>
<td>Patients feel reassured that “someone is monitoring”</td>
<td>—</td>
</tr>
<tr>
<td>Patient-health care institution (4)</td>
<td>Institutional trust</td>
<td>Patients feel reassured that “someone is monitoring”</td>
<td>Institutional risk</td>
</tr>
<tr>
<td>Patient-telehealth technology (5)</td>
<td>Human-technology trust</td>
<td>Patients’ trust in technology and telehealth</td>
<td>Patients’ distrust of or discomfort with technology, Patients’ experiences of uncertainty</td>
</tr>
<tr>
<td>Health care professional-telehealth technology (6)</td>
<td>Human-technology trust</td>
<td>Clinicians’ trust in technologically generated data</td>
<td>Clinicians’ concerns regarding the limitations of technology-mediated interactions, Clinicians’ experiences of uncertainty</td>
</tr>
<tr>
<td>Patient-telehealth technology-health care professional (1 and 2)</td>
<td>Interpersonal trust</td>
<td>Mutually trusted data support patient-clinician partnership</td>
<td>—</td>
</tr>
</tbody>
</table>

*No themes emerged from the data in this category.*
Telehealth as Supplementary Rather Than Substitutive

Although some researchers [15, 56] have raised concerns that depersonalization could occur and trust relationships would be damaged with the adoption of telemedicine—the “transformation of the fiduciary relationship into a more contractual or quasi-contractual relationship” [15]—we found the reverse to be true. Increased contact with the health care system, albeit remotely via patients’ participation in uploading their BP readings and teleconsultations, seems to make the health care provider more continuously present to the patient than before. Patients felt that they were more closely followed up with and were more aware of health care providers as partners in managing their health. However, as a physician observed, the trust relationship needed to be properly established in person before telemedicine was introduced. This is consistent with observations by van Middelaar et al [57] and others [58] that an eHealth intervention is more readily trusted when eHealth is combined with at least an initial in-person interaction with a trusted offline entity at the outset.

Our study’s findings suggest that telemedicine is not necessarily detrimental to the human touch or bedside manner of the physician, as some critics fear. In fact, the generally positive feedback indicates that telehealth has much potential to supplement (though not totally replace) face-to-face health care by extending the care and attention given by health care providers beyond the physical boundaries of the clinic.

Remote BP Monitoring Creates Ubiquitous and Continual Presence of the Health Care Provider

From a patient perspective, the extension of care beyond the walls of the polyclinic via the remote BP monitor and teleconsultations blurs the boundaries of care, blurring the world of clinical treatment with the intimacy of patients’ daily lives. The elements that constituted the telehealth program—the physical presence of the telehealth equipment, phone calls from the health care provider, and patients’ action of uploading their readings weekly—engaged patients in their own care and transformed the health care provider-patient relationship from a predominantly episodic one bounded by time (of the patient’s appointment) and space (the venue of the polyclinic) to one that formed part of the fabric of their ordinary lives, leading to a greater sense of trust in the health care provider. Thus, patients’ experience of health care shifted from being periodic and transactional to an ongoing, continual relationship with the health care provider virtually present in their homes by means of remote BP monitoring and telephone consultations.

Over time, because of the technologically mediated interactions with the health care system and telephone consultations, patients felt greater familiarity with the health care professionals. As a result, some patients appear to have more readily accepted physicians’ advice on medication adjustment.

Trust transference from the in-person context to the technological one is likely to have played a key role in the first 2 themes as all patients enrolled in the study had previously engaged in face-to-face encounters with clinicians at this health care facility, although not necessarily with the same clinician. Several studies have highlighted the existence of trust transfer in eHealth, “from brick to click.” Van Velsen et al [59] found that, for patients, trust in the care organization was conceptually different from trust in the care team and trust in the treatment but that trust in the care team and trust in the treatment affected trust in the technology. Our findings also support those of Meng et al [60] and Pavlova Miller [40], who found that trust in offline health services was positively associated with trust in web-based health services.

Selective Uploading and Naïve Trust in Systems

Interestingly, the health care professionals interviewed tended to perceive the technologically mediated readings as more reliable than manually recorded ones, on the assumption that patients “can’t cheat” because the BP reading is automatically uploaded to the system. However, some patients sought to control or manipulate the BP readings uploaded so that only desirable readings would be sent to the system. This implicit trust in technology observed from the health care professionals was at odds with the discovery that patients may in some cases modify what is recorded to selectively present their biodata. Choosing to upload only desirable readings may be attributable to a form of social desirability manifesting itself as the desire to present oneself as a “good patient” [61-63]. Therefore, it is pertinent for health care providers using telehealth to be aware of the role human factors and motivations play in patient behavior. They will in this way avoid assuming that data uploaded by patients and produced in the context of patients’ lives occurs in a completely objective environment devoid of subjective and extraneous influences.

Managing Uncertainty and Risk

Our key themes reflected various aspects of both trust and uncertainty associated with telehealth technologies. Uncertainty is antithetical to trust; trust and uncertainty have been described as “a pair of opposing forces shaping relationships” [64] in dialectical tension. To decrease uncertainty is to help foster conditions necessary for trust.

Our study showed that uncertainty was often present in the telemedicine interactions. For patients, there was uncertainty about where the biodata would be stored, lack of feedback on whether the BP measurement uploading process was successfully completed, if and when the submitted data were being monitored by the health care provider, and the inability to clarify doubts about their medical conditions or medications.

The interviewees who distrusted technology became anxious about their ability to successfully upload their readings (ie, low technological self-efficacy) and tended to be more likely to drop out of the program. This is in line with previous findings that users’ postadoptive behaviors are affected by trust in the technology and that technological self-efficacy may be a mediating variable for trust in technology [65]. Greenhalgh [66] points out that technological innovations often fail because “the patient in the guideline does not correspond to the patient in the bed”—telemedicine initiatives often envision an empowered, self-motivated patient who understands, trusts, and happily uses the technology. Our study found this to be true of many patients, but others, especially certain patient interviewees aged >50 years, were distrustful of or apprehensive about technology. For
this program to be effective and sustainable in the long run, additional efforts should be made to reduce uncertainty and raise the level of comfort with the telehealth technology for such segments of the population.

For clinicians, uncertainty arose from the lack of visibility of a patient’s actual health status because of the limited data circumscribed by the properties of the technology used. To increase the sustainability and acceptability of this program, it would be helpful to increase the trust of all human stakeholders by reducing the uncertainties faced by patients and staff. Design choices such as having clear feedback from the device when the reading has been correctly uploaded and explicit indications of how often the patients’ uploaded data will be reviewed could go some way toward reducing uncertainty for patients. Reducing uncertainty would make it easier for staff and patients to trust the technology and for patients to increase their confidence in the telehealth program as well as in the health care provider.

Managing institutional risk and the uncertainties it causes to organizational stakeholders is also important for any telehealth program. Support can be provided through clear guidelines, as was done in this case, to limit uncertainties about legal liabilities, reputational risk, and other repercussions on health care professionals arising from possible failures in the telehealth program.

**Limitations**

This study undertook an exploration of user views during the implementation process, supplementing the findings from a larger mixed methods study. Slightly more than half (7/13, 54%) of our interviewees were aged ≥50 years, reflecting population prevalence as hypertension is much more common in this age group [67]. More work is needed to understand the needs of younger patients with early onset hypertension as they are likely to engage differently with telehealth.

As this study interviewed patients enrolled in a remote BP monitoring program, some selection bias is to be expected as potential participants who are deeply averse to technology would have declined to participate. Future studies should specifically seek out views of segments of the population who are less inclined toward telehealth to elucidate their concerns.

**Conclusions**

In our study, telemedicine was used to complement existing face-to-face care by reducing physical clinic visits while increasing the monitoring of patients’ health via technology-enhanced remote BP monitoring. Our findings elicited aspects of patient trust in health care providers (as individuals and as an institution) as well as in the telehealth technology and found elements that encouraged or hindered the building of trust in an existing patient-health care provider relationship. Generally, patients greatly valued the closer follow-up, which was also deemed more personal, although a few refused to relinquish or reduce in-person follow-up visits. Future work could investigate the possibilities of telemedicine to extend the human touch in remote medical care rather than substitute it. Well-designed telehealth interventions can remotely extend the presence and sense of closeness of the health care provider and, thus, increase quality of care without detriment to productivity, resulting in stronger partnerships with patients in managing their health.

Other aspects of the trust relationships warrant further research, such as how the affordances and design of a specific telehealth intervention affect perception and trust, the impact of telehealth on interprofessional trust relationships, and how telehealth affects the health care provider-patient relationship in the absence of a previous offline relationship.

Attitudes toward new technologies are often mixed, with some stakeholders enthusiastic about the novelty and others critical, skeptical, reluctant, or even hostile [9,10]. However, for telemedicine to work well, trust is crucial [68]. Exploring the impact of a telehealth intervention on trust relationships helps shape future developments of similar projects with a view to maximize benefits, avoid pitfalls, and enhance patient relationships with their health care providers. It is hoped that this exploration of the dimensions of trust in a telehealth program will assist designers and implementers of telehealth as well as health care researchers in taking cognizance of the role of trust and other human factors in their telehealth program development and its implementation.
Conflicts of Interest

None declared.

Multimedia Appendix 1

Patient and staff interview topic guide.

[PDF File (Adobe PDF File), 77 KB - humanfactors_v10i1e36072_app1.pdf]

References


Abbreviations

BP: blood pressure

Edited by A Kushniruk; submitted 14.01.22; peer-reviewed by L van Velsen, D Madhusudhan; comments to author 26.03.22; revised version received 10.08.22; accepted 06.10.22; published 05.01.23.

Please cite as:

Chew E, Teo SH, Tang WE, Ng DWL, Koh GCH, Teo VHY
Trust and Uncertainty in the Implementation of a Pilot Remote Blood Pressure Monitoring Program in Primary Care: Qualitative Study of Patient and Health Care Professional Views
JMIR Hum Factors 2023;10:e36072
URL: https://humanfactors.jmir.org/2023/1/e36072
doi: 10.2196/36072
PMID: 36602847
Characterizing the Gaps Between Best-Practice Implementation Strategies and Real-world Implementation: Qualitative Study Among Family Physicians Who Engaged With Audit and Feedback Reports

Geneviève Rouleau1,2, RN, PhD; Catherine Reis1, BA; Noah Ivers1,3,4, MD, PhD; Laura Desveaux1,3,5, MScPT, PhD

1Institute for Health System Solutions and Virtual Care, Women’s College Hospital, Toronto, ON, Canada
2Research Chair in Innovative Nursing Practices, Centre de recherche du Centre hospitalier de l’Université de Montréal, Montreal, QC, Canada
3Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada
4Department of Family and Community Medicine, University of Toronto, Toronto, ON, Canada
5Institute for Better Health, Trillium Health Partners, Mississauga, ON, Canada

Abstract

Background: In Ontario, Canada, a government agency known as Ontario Health is responsible for making audit and feedback reports available to all family physicians to encourage ongoing quality improvement. The confidential report provides summary data on 3 key areas of practice: safe prescription, cancer screening, and diabetes management.

Objective: This report was redesigned to improve its usability in line with evidence. The objective of this study was to explore how the redesign was perceived, with an emphasis on recipients’ understanding of the report and their engagement with it.

Methods: We conducted qualitative semistructured interviews with family physicians who had experience with both versions of the report recruited through purposeful and snowball sampling. We analyzed the transcripts following an emergent and iterative approach.

Results: Saturation was reached after 17 family physicians participated. In total, 2 key themes emerged as factors that affected the perceived usability of the report: alignment between the report and the recipients’ expectations and capacity to engage in quality improvement. Family physicians expected the report and its quality indicators to reflect best practices and to be valid and accurate. They also expected the report to offer feedback on the clinical activities they perceived to be within their control to change. Furthermore, family physicians expected the goal of the report to be aligned with their perspective on feasible quality improvement activities. Most of these expectations were not met, limiting the perceived usability of the report. The capacity to engage with audit and feedback was hindered by several organizational and physician-level barriers, including the lack of fit with the existing workflow, competing priorities, time constraints, and insufficient skills for bridging the gaps between their data and the corresponding desired actions.

Conclusions: Despite recognized improvements in the design of the report to better align with best practices, it was not perceived as highly usable. Improvements in the presentation of the data could not overcome misalignment with family physicians’ expectations or the limited capacity to engage with the report. Integrating iterative evaluations informed by user-centered design can complement evidence-based guidance for implementation strategies. Creating a space for bringing together audit and feedback designers and recipients may help improve usability and effectiveness.

(JMIR Hum Factors 2023;10:e38736) doi:10.2196/38736
KEYWORDS
Audit and feedback; family physicians; primary care; qualitative

Introduction

Background
Audit and feedback (A&F) is a quality improvement (QI) intervention that involves the collection and analysis of population- or practice-level data (audit) and the provision and delivery of clinical performance summaries (feedback) [1,2]. A&F is widely used across health care settings [3-5] by a variety of stakeholders, both to increase accountability and to improve quality of care [6]. A wide range of behaviors may be targeted, including but not limited to laboratory testing and transfusion ordering [3], adherence to clinical guidelines, and prescription [7]. Many factors influence A&F effectiveness, including the characteristics of the targeted behavior, recipients (eg, their skills and capabilities), A&F itself (eg, feedback display and delivery), and context [2,8-10]. Some targeted behaviors and contexts may be more amenable to A&F. However, all health professionals have the potential to benefit from A&F, underscoring the need to better understand whether and how to align the nature of A&F itself with the characteristics of the recipients.

Evidence indicates that the greatest effects of A&F may be achieved by optimizing the frequency of the feedback, the format in which it is delivered (verbal, written, or both), the use of visual display, the provider of the feedback (eg, a supervisor or colleague), the content of the feedback, the provision of explicit goals, and action plans [8,10-12]. Regardless of the design choices for the intervention, to change clinicians’ practice and, subsequently, patient outcomes, clinicians must first engage with A&F and then act upon the messages within.

Not all evidence-informed best practices of A&F are easily operationalized, and some may have been designed in a variety of ways (eg, color choices, positioning or size of information, or specific word choices used to describe performance). The extent to which this affects whether A&F recipients engage with it is uncertain. To address this gap, we undertook a qualitative evaluation of the redesigned A&F report. We used the COREQ (Consolidated Criteria for Reporting Qualitative Research) [13] guidelines for reporting the qualitative process.

Objectives
The initial objective of this study was to evaluate whether and how the redesign improved the usability and perceived effectiveness of the report. Early interviews challenged our underlying assumption that recipients were meaningfully engaging with the original report. We then shifted our objective to exploring the perceived usability of the report in general, with an emphasis on recipients’ understanding of the report and their engagement with it. We sought to generate recommendations on how A&F designers can work with A&F recipients to successfully operationalize best practices in the real world.

Methods

Ethics Approval
This project was formally reviewed by the institutional authorities of the St. Michael’s Hospital Research Ethics Board (16-076). The Women’s College Hospital Research Ethics Board performed an administrative review of the study (2016-0136-E) and granted the research team an exemption from Research Ethics Board review for this study. Verbal consent to participate in the study was obtained by the interviewer (CR).

Study Design
We conducted qualitative semistructured interviews to understand how family physicians perceived and engaged with the redesigned A&F report. We used the COREQ (Consolidated Criteria for Reporting Qualitative Research) [13] guidelines for reporting the qualitative process.

Context and Setting
Ontario Health, an agency created by the Government of Ontario with a mandate to connect and coordinate the health care system, offers a range of resources to support health professionals in providing better care. This includes providing physicians and family health teams with information about how their practices compare with those of other physicians across the province via the MyPractice Primary Care report. In Ontario, primary care is delivered mainly by family physicians. The provincial health insurance plan, funded by the Government of Ontario, pays for all physician visits, tests, and prescription medications measured in the MyPractice Primary Care report.

Intervention—MyPractice Primary Care Report
The MyPractice Primary Care report was initially developed by Ontario Health in partnership with the Association of Family Health Teams of Ontario, the Association of Ontario Health Centres, and the Ontario College of Family Physicians. The stakeholders involved in developing the original report were members of regulatory organizations, working primarily at the system level, knowledgeable of population-related data, and familiar with these types of initiatives. Physicians who were members of the cited organizations were not necessarily the end users of this report.

At the time of the study, administrative data sources were used to assess a series of quality indicators: safe prescription (eg, opioid and benzodiazepine prescription rates), cancer screening (eg, percentage of patients with up-to-date cancer screening tests for cervical, breast, and colon cancer), diabetes management (eg, percentage of patients with diabetes who had had ≥2 HbA1c tests within the past 12 months, who had diabetes and were aged >65 years and had an active statin prescription, and who had had a retinopathy screening test within the previous year), and health service use (eg, emergency department visits, hospital admissions, and readmissions [by condition]).
Administrative data also encompassed clinical (chronic disease) and demographic (age and income) information on the patient population. Aggregate-level data were presented for each of the indicators, covering the previous 12 months of clinical practice. Practice improvement ideas specific to each of the topics were included to support recipients in taking action.

Family physicians in Ontario must sign up to receive this report. The original reports were designed without formal user testing. We developed a new prototype based on the original report, with attention to best practices [1]. We then refined the prototype with 16 naive users (family physicians who had not signed up for the MyPractice report) by observing them interact with the A&F report (usability testing) with the aim of improving usability. Usability sessions involved observing participants to determine how they were navigating and understanding the A&F prototype. Participants were asked whether they were unsure about or had trouble understanding any aspects of the report and whether anything might be missing that could be helpful. The findings led to changes in the graphic design, a revised visual summary of performance compared with peers on the quality indicators, and an attempt to more clearly connect the aggregated quality indicators with suggested actions for improvement. The final product of these usability sessions was the redesigned report. Versions of the original and redesigned reports can be found in Multimedia Appendix 1. At the time of the study, to access the report, family physicians had to log in to a password-protected website. Starting in May 2017, the report was emailed to the participating family physicians. The overall development and evaluation processes of the A&F reports is presented in Figure 1.

**Figure 1.** Overall development and evaluation processes of the A&F reports.

---

**Participants**

Eligible participants were Ontario family physicians who were registered to receive the redesigned report following its release in May 2017 and had experience with the original version. These individuals were contacted via email by Ontario Health and invited to participate in a one-time interview with a member of the research team. A CAD $100 (US $73.36) honorarium in the form of a gift card was offered. Recruitment continued until data saturation was reached. A convenience sampling approach was used whereby an email outlining the study was sent to those who registered to receive the report. Those who received the email were also encouraged to share the study information with their colleagues who had also viewed the MyPractice report (i.e., snowball sampling).

**Data Collection**

Semistructured interviews were conducted over the phone by a member of the research team (CR; see Multimedia Appendix 2 for the interview guide). The interviews were audio recorded and transcribed verbatim by an independent third party. While viewing their confidential report, participants were asked about their overall impressions of it, whether they felt it was easy to navigate, and what parts of it they found most useful and why. Participants were also asked to describe what actions (if any) they took following their review of the report (e.g., conducted a chart review to determine who among their patients with diabetes was due for an HbA1c test) and what features of the report informed those actions.

**Data Analysis**

The qualitative analysis followed an emergent and iterative approach. In total, 3 members of the research team (CR, NK, and BB) independently coded a first transcript; all interviews were double coded. They then met to compare the interpretation of the targeted quotes and revise the codebook. A peer debriefing was conducted—preliminary findings were discussed in a meeting with senior investigators (LD and NI) who have both conducted multiple previous qualitative studies involving A&F. The team concluded that the data were pointing to broader questions related to the perceived usability of the reports to support QI in practice. At this stage, the team reanalyzed the data using a more inductive lens. LD coded 2 transcripts to become immersed in the data. The focus of the data analysis shifted to an emphasis on recipients’ understanding of the A&F reports and their engagement with them. A conventional content analysis was performed following an inductive approach [14]. The codes and categories were iteratively revised throughout this second stage of data analysis. CR then met with another member of the research team (GR) to discuss and refine the categorization of codes and establish themes. In total, 4 members of the research team (CR, GR, LD, and NI) met to further refine the themes, which were then finalized by all authors. CR
maintained a consistent audit trail of the codebook throughout the 4 stages of data analysis.

Retrospectively (ie, once the data analysis was completed), we mapped the specific redesign elements to the corresponding A&F best practice [1] they were intended to operationalize. We further mapped these applied recommendations to the corresponding theoretical constructs as outlined in Clinical Performance Feedback Intervention Theory (CP-FIT). The final themes were described with these elements to explore areas of success and failure to generate insights to optimize the real-world implementation of best-practice guidance.

Results

Participant Characteristics and General Interaction With A&F
A total of 17 family physicians participated in the interviews (n=8, 47% female participants and n=9, 53% male participants) lasting from 15 to 60 minutes. In total, 41% (7/17) of the participants had between 1 and 10 years of practice experience, and the remaining 59% (10/17) had >20 years in practice. Most (10/17, 59%) worked as part of a multidisciplinary family health team. Participants appreciated certain design elements such as the targeted use of color and emphasis on the number of eligible patients for a specific action as they facilitated review and interpretation of the data. This helped them better understand the data. However, family physicians described challenges in identifying actions to take in response to the data that undermined the overall utility of the report. Factors that affected the perceived usability of the report can be summarized in two key themes: (1) alignment between the report and recipients’ expectations and (2) capacity to engage with QI.

Theme 1: Alignment Between the Report and Recipients’ Expectations Affects Usability

Overview of Theme 1
Family physicians described their expectations of the feedback report related to the quality indicators and data presented. First, they expected the report and its indicators to reflect best practices. Second, they expected the quality indicators to be valid and accurate. Third, family physicians expected the report to offer feedback on the clinical activities that they perceived to be within their control to change. Finally, family physicians expected the goal of the report to be aligned with their perspectives on QI. When these expectations were not met, the perceived usability of the report was low. Quotes supporting the subthemes of theme 1 are presented in Textbox 1.
Subtheme 1.1: quality indicators must reflect best practices

- “You have to make sure...numbers are important, but the number has to reflect purpose. When you give a precise number for something that’s meaningless, you have precision of something which isn’t going to motivate.” [Participant 12]

- “I’m below average for LDL testing for diabetic patients, mostly because it looks like they’re looking at me doing annual LDL testing. Personally, I think the evidence points to not actually doing this on a routine basis. And I’m at average or I’m above average with respect to statin prescriptions for those diabetic patients. So, I think that kind of fits more of what we’re trying to get at, rather than the LDL testing...I think that testing LDL doesn’t necessarily help outcomes for my patients.” [Participant 6]

Subtheme 1.2: quality indicators must be perceived as being valid and accurate

- “The last line that goes over the demographics is really interesting. I seem to have more of the older-age, geriatric practice and it’s kind of nice to see that because I think that influences referrals and how many times people go to the Emerg as opposed to practices that may have a much younger population. So it’s really nice that I think it acknowledges the demographic of your practice.” [Participant 11]

- “It gives a whole bunch of people that are not up to date with hemoglobin A1c testing, but it’s incorrect data. It says that most of our diabetics, I think our line is 13%, which is incorrect. So, all this stuff is not useful for me.” [Participant 8]

- “Knowing that the data is not accurate, because it’s based on [public databases]...I have less buy-in that the data actually reflects my real practice. Simply because there is no way for me to feed back to the system, either through [this report or others], to say that on this particular patient on this data point you don’t have it right.” [Participant 9]

- “Nice to have a reference of how we do compared to the rest of the province. That’s part of the thing that’s valuable, it gives us an indicator, when you get a comparator of how the provincial average is.” [Participant 15]

Subtheme 1.3: quality indicators are expected to be actionable and within physicians’ control

- “I think it’s in my mind more. For instance, the retinal testing I was slightly below so it was just on my mind when I’m doing my diabetic checks...It primes me to do that.” [Participant 3]

- “I get a little irritated...I mean, if you’re doing everything you can, it’s a little frustrating, because you wonder what you can do more. With these numbers, with the A1C, I see most of my diabetics every three months, so I’m thinking, well, why is that going down? Also, with the retinal scan. I mean, you have to ask them if they go to the eye doctor and they say, yes, but clearly, according to this, it’s going below the average, which means...It can be good, but it’s also frustrating, because a lot of times this is stuff out of your control...I think we have to adapt our indicator to remember that people will make their own decisions and we don’t have control.” [Participant 14]

- “I don’t think that there are many things in my control to change those numbers and so going on again and again has felt kind of like a waste of time because I’m quite sure that nothing will be different.” [Participant 16]

Subtheme 1.4: alignment of the goal of the report with how physicians approach quality improvements

- “I think the question I have...is what you would like physicians in general to do with the report? Because it’s all nice to give people information but if there is no clear direction about what they should do with it and how they could integrate it easily into their day-to-day use of their EMR [electronic medical record] or of their function in the office.” [Participant 9]

- “So if you’re using [data] as a guide to help physicians improve their practice that’s one thing, but if you’re using it to evaluate physicians, I think the data is just not good enough for that.” [Participant 7]

- “There’s great cancer screening, for sure, in terms of seeing where I’m at with that, seeing if, we do invest quite a bit of our staff time and energy into calling and mailing patients and reminding of that stuff. And so, to see that that’s paying off and that we’re not doing all that work and still below average or something. That’s very validating.” [Participant 5]
over objective numbers. When the data are perceived as not useful, this negatively affects physician buy-in.

**Subtheme 1.3: Quality Indicators Are Expected to Be Actionable and Within Physicians’ Control**

When reviewing feedback in their report, physicians interpreted their current performance as reflective of either action or inaction on their part or that of their patients. Family physicians expected the report to offer feedback on the clinical activities that they perceived to be within their control to change. When family physicians determined that an indicator within the report reflected activity beyond their control, they determined that the indicator was irrelevant to their practice and did not expect to see improvement over time. Being aware of and in agreement with an area of practice requiring improvement can prime action. A major limitation of the quality indicators was the inability to capture the shared decision-making process and the person-centered approach. A physician can offer guidance and direction, but it is ultimately the patient who takes action either completing a test or receiving a treatment. Physicians expressed some frustration as the indicators were not reflective of this shared responsibility.

**Subtheme 1.4: Alignment of the Goal of the Report With How Physicians Approach QI**

Finally, family physicians were unclear as to the goal of the report and expressed a need for clearer direction or an explicit target to support action. They expected the report to be aligned with their perspective on QI: supporting point-of-care decisions by identifying areas of improvement, offering clear guidance on how to improve performance, and identifying specific targets in line with desired actions. The report was not perceived as a means of evaluating physicians' performance as the data were not “good enough” to support this type of evaluation.

However, family physicians appreciated the opportunity to see change, specifically improvement in their performance following concrete efforts to improve.

**Theme 2: Capacity to Engage With QI Affects Usability**

**Overview of Theme 2**

Even when family physicians agreed that reviewing their performance data was an important part of their professional role, they described several barriers to engaging with the report. System-level conditions (e.g., time and resources) as well as work-related conditions (e.g., workload and competing priorities) affected different stages of the QI process, including accessing the data, interpreting the data, and action planning. Quotes supporting the subthemes of theme 2 are presented in Textbox 2.

Textbox 2. Quotes supporting the subthemes of theme 2 (capacity to engage with quality improvement affects usability).

---

**Subtheme 2.1: hard to fit A&F into the workflow and resource constraints**

- “We balance prevention with everything else that we do...if we followed all the good evidence in terms of prevention, and not just the things that...are in these reports...Those people, if we do what the evidence says we should do for prevention in the top ten chronic diseases, there is no time to do all the other stuff. You have to be reasonable about how we put our efforts. We could get these indicators up a lot higher, but people would be dying. It’s good that we are doing this. I’m not saying there is anything wrong with that. But the context is, this is only a tiny part of what we do. You have to look at your resources.” [Participant 13]

- “It just is one less step because if I see that I have 27 patients not tested for diabetes, I have to dangle into my EMR and do the search myself. So it’s extra searching and busy day it might not become the top of my list. But if it’s right there for me then I’m going to be more likely to follow up on that.” [Participant 3]

**Subtheme 2.2: insufficient knowledge and skills to interpret the data**

- “I’m just not sure how to interpret it. We’ll say, for example, total Emergency room visits. It tells me my practice, unadjusted, is 810 visits per 1,000 patients. Then, in the next column over, it does a risk adjustment and downgrades it to 504. I presume what that means, but I’m not entirely clear, is that my practice may be more complicated or have more comorbidities, so my number actually isn’t as bad as 810, that it’s gone down to 504 to account for that. But, again, I’m curious about that. I don’t know, does that mean I can take away from that, that I have a more complicated practice than average?” [Participant 1]

- “It’s nice to compare myself to other people but I guess what I look at, is going oh I’m doing better than everybody pretty much on everything except, you know, so now what. Just because I’m better than everybody else does that mean I’m good enough? I don’t know what that means...” [Participant 9]

**Subtheme 2.3: lack of guidance on how to prompt actions**

- “You need to be able to see how you’re doing on the big scope of things, yeah, but you need to be able to thin it down to the individual patients that make up the bigger picture. That’s what spurs the action, to identify who they are.” [Participant 2]

- “It would be nice, just to compare yourself to other people in our immediate group. I think that probably has a little more educational kind of component to it if someone is doing a lot better with something than everybody else, hey, maybe they’re doing something we can copy or emulate. It’s hard to copy people you don’t know and don’t work with and never see. So, it’s easier to engage the change idea stuff if you’ve got someone on the ground, close to you, that’s doing something different.” [Participant 2]

- “I think it’s nice to see the trend but at the same time how do we act on it now? And that’s what kind of deterred me from moving forward and using it more often. So I think our EMR would...and when we do a search we actually shoot out here are the patients who are overdue and then our nurses and team try to call those patients or keep it in the back of our minds. I think the summary is super nice to look at, out of interest, but again it’s not helping at a patient-specific level...” [Participant 11]
**Subtheme 2.1: Hard to Fit A&F Into Workflow and Resource Constraints**

Competing priorities were a reality for family physicians—they had a heavy workload of clinical tasks each day. They also had to navigate through different duties and roles as educators, leaders, and managers. Some of them reported balancing their time between preventing and treating diseases, which influenced what activities they prioritized. They had to weigh carefully how additional QI processes in response to A&F might fit into their workflow. Physicians highlighted that accessing their data, which means searching their patient records, was a time-consuming process that was hard to integrate into the existing workflow.

**Subtheme 2.2: Insufficient Knowledge and Skills to Interpret the Data**

Physicians struggled to interpret some aspects of the report and questioned the meaning of their data. Some participants clearly mentioned not knowing what to do with the aggregated practice-level numbers. A participant suggested that discussing the content of the report with someone they trusted would be helpful.

**Subtheme 2.3: Lack of Guidance on How to Prompt Actions**

Regarding action planning, physicians perceived the report as unactionable as (1) it was not perfectly up to date and (2) the aggregated nature of the data could not easily be translated into clinical actions without additional support. These challenges were not at all influenced by the visual nature of the report redesign. However, some participants appreciated the value of having their data compared with those of the rest of the province. These comparisons helped them evaluate their performance and made them aware of areas of improvement in their practice. However, the evidence-to-practice gap remained, and they did not know how to use the data to change their practice.

**Alignment Between Best-Practice Implementation Strategies and Actual Implementation**

Supported by the content of Multimedia Appendix 3, we present the success and failure of our study. What seems to have worked (ie, alignment between recommendations, target of the redesign, and participants’ perspectives) was choosing comparators that reinforce desired behavior change and linking the visual display and summary message. Physicians valued the comparisons and found them helpful in pinpointing areas of improvement. However, they struggled on how to interpret the meaning of these comparators and how to consequently change their practice. Few participants commented on the visual of the feedback display, but those who did found it “nice.” This highlights a need to further understand and evolve the way recommending actions that can improve and are under the recipient’s control is operationalized. In the redesigned report, providing brief information regarding the importance of action on each given indicator and highlighting an absolute number of patients that appeared to require action for a given indicator did not sufficiently help recipients understand how to act. On the one hand, when physicians were in agreement with an area of practice under their control requiring improvement, it seemed to be a motivation to take action. In contrast, physicians reported that some of the indicators were beyond their control and reflected elements of care that relied on shared decision-making and patient action. Efforts in addressing the credibility of information and in recommending actions consistent with established goals and priorities were unsuccessful in promoting physicians’ engagement with the A&F report. Participants did not trust the source of the data and perceived that the information lacked relevance to their practice, was not a priority, and was not motivating. Furthermore, they did not understand how the aggregated nature of the data could translate into a way that informed clinical actions. Several physicians cited the need for counterinterventions, such as peer discussion point-of-care reminders and support with action planning, to support meaningful engagement and subsequent practice change.

**Discussion**

**Principal Findings**

The findings suggest that, although the redesign did improve the “look and feel” of the A&F, it was not sufficient to drive practice changes in response to the data. Even if they were unconvinced that the indicators were the right targets for action, some physicians became newly self-aware of the gaps in their practice. Although this awareness may be a trigger for initiating professional behavior change processes, the desired actions for QI were unfortunately perceived as uncertain or unfeasible. The capacity to engage with the MyPractice report was hindered by several organizational and physician-level barriers, including the lack of fit with existing workflows, competing priorities, time constraints, and insufficient guidance and skills regarding how to interpret the data and bridge the gaps between their data and the corresponding actions.

**Lessons Learned**

**Overview**

To understand the implications of the findings of this work, we applied three perspectives: (1) the recommendations by Brehaut et al [1] derived from stakeholder interviews to identify what elements of best practice need to be included in A&F interventions to improve their effectiveness; (2) the CP-FIT [2] derived from a qualitative systematic review and meta-synthesis to provide insights on how users typically progress through an A&F cycle and then help understand users based on the elements relevant to the A&F cycle (what do we need to know about users that is more important to A&F); and (3) user-centered design principles, including empathy with the end users’ goals and an understanding of their context, to operationalize those elements [12,15]. Our findings indicate the potential for integrating these perspectives into a single lens when developing and refining A&F.

**Understanding Users, A&F Interventions, and Contextual Elements as a Whole**

In the refinement steps for the feedback process anchored in a user-centered design approach, Landis-Lewis et al [12] show how refining measures, data, and display can be embedded in the development and refinement step of an A&F prototype. In
a learning health system approach, implementation strategy design should be iterative, informed by the ongoing collection of real-world data. Our findings highlight the importance of testing implementation strategies in context. They also echo 3 variables as proposed in the CP-FIT model [2] that influence the feedback cycle: recipient, feedback, and context variables. Figure 2 illustrates this approach to iterative A&F development following user-centered design principles that considers end users, contextual elements, and A&F intervention components.

Figure 2. User-centered approach to develop and refine the feedback report based on a synergic understanding of users, contextual elements, and the audit and feedback (A&F) intervention (informed by Landis-Lewis et al study [12]). CP-FIT: Clinical Performance Feedback Intervention Theory.

Understanding the Feedback Recipients (End Users) From the Very Beginnings of Designing A&F and Seeking Out a Variety of Perspectives

It would have been appropriate to engage A&F recipients [12,15,16] from the very beginnings of designing A&F, which was not the case in this initiative. The project stakeholders (eg, Ontario Health and Association of Family Health Teams of Ontario) who designed the initial report are not the ones using the A&F report in their practice. To discover different users’ perspectives when designing the A&F intervention, it would have been useful to seek out A&F recipients with various characteristics, such as high, moderate, and low degree of exposure to A&F interventions; degree of agreement with those initiatives; and high and low performances. Although the usability sessions were conducted with naïve users, they may have focused too narrowly on the intervention elements rather than on the intervention goals. In contrast, Cooke et al [17,18] illustrated a process in which A&F report designers and physicians (end users) collaborated to design and implement A&F. The physicians identified key clinical questions, made individualized A&F reports, and developed a plan for change through participation in a group feedback session. By incorporating end-user feedback into the design of A&F, user-centered design helps ensure that the design of A&F reports is functional; can support end-user needs and goals; and, ultimately, positively influences clinical practice [12,19]. If we had used this approach, it is possible that the A&F report and recipients’ expectations would have been more aligned.

Comparison With Prior Work

It is possible that overarching best practices for designing and implementing A&F [1,20,21] should be seen as hierarchical—some may matter more than others. For example, our study shows that, even if the “design” features of feedback display can all be addressed (eg, provide feedback in more than one way, such as presenting key messages both textually and numerically), if the focus of the A&F is not aligned with recipients’ goals and the audit itself is perceived as lacking validity, accuracy, and credibility; is poorly aligned with physicians’ priorities or not readily actionable; and is not under their control, then the intervention will not achieve its potential to improve quality. A prioritization exercise among 61 A&F stakeholders to identify the top 50 “priority” foci for the A&F research agenda [22] produced understandable variability; however, 50% of the participants identified hypotheses relating to the factors that we identified as relevant to engagement. These include testing the impact of a trusted source (“trustworthiness/credibility”), recipients being involved in the development of the feedback intervention (decision processes or conceptual model), a foundation of good-quality evidence (“trustworthiness/credibility”), and the behavior being under the control of the recipient (“self-efficacy/control”). In line with our findings, the form of the A&F reports was not extensively discussed by the participants, which led us to believe that this was not a priority for improving the effectiveness of the A&F report.

The importance and relevance of feedback goals are key variables that affect recipients’ acceptance and their intention to change their behaviors—2 key elements of the feedback cycle as outlined in the CP-FIT [2]. Family physicians in this study wanted clearer direction of what to do with the report but also clarity on the purpose and meaning of the entire A&F initiative (ie, evaluating and measuring physicians’ performance vs improving practice). Recommendations regarding the nature of desired actions further specify the need for alignment with established goals and priorities [1], which may be enhanced by including an exemplar action plan that could be adopted in response to the A&F [20]. Some family physicians thought that the quality indicators did not fairly reflect their practice and attributed the data to patient behaviors (eg, screening). In this case, physicians felt that the data represented activities beyond their control, highlighting the importance of controllability.
which can negatively affect the acceptance of the report [2]. Physicians noted that best-practice elements of care, specifically patient-centeredness and shared decision-making, were not reflected in their data. Consequently, physicians felt judged for their performance based on data for which they were not entirely responsible, causing frustration. Other studies have also highlighted the need for quality indicators to reflect the important role of patient choices [6,23] as well as measures representing patients’ perspectives on care, clinical quality, and general quality of care from a broader perspective [24].

Family physicians’ views and QI knowledge and skills (or lack thereof) influence how they interact with A&F [2,23], highlighting the need for cointerventions. This corresponds to the recipient variable in the CP-FIT, specifically to Knowledge and skills in QI. In this study, physicians did not know how to act upon their data even though the redesigned report attempted to more closely connect the data with recommended actions (ie, “change ideas”). Educational strategies delivered alongside A&F have been effective in supporting improved adherence to guidelines [25], reducing the rate of cesarean delivery [26] and antibiotic prescription [27]. In these studies, strategies were operationalized in different ways, such as a 1-hour group session [25], quarterly educational outreach visits conducted by external facilitators [26], and 2 sessions of voluntary continuing medical education in addition to educational materials [27]. Training-based interventions effectively build skills [28] and improve communication skills [29], whereas an emphasis on data interpretation and action planning is likely to positively influence practice change [2,30]. Considering that passive feedback delivery (ie, written and delivered through email) might have played a role in physicians’ engagement, adding active interactions (eg, peer discussion or other social interactions) throughout the feedback cycle is likely necessary [18,23]. However, social interaction alone is likely to be ineffective if it does not incorporate a component of prompting actions, which may include asking targeted and reflective questions about what can change [18] or highlighting and sharing the actions of high performers [31].

These insights highlight several areas of focus as the science and implementation of A&F moves forward (Textbox 3).

**Textbox 3. Areas of focus highlighted by insights from this study.**

**Focus areas**

- Using a user-centered design approach that considers the characteristics of and interactions between the users, their context, and the characteristics of the A&F interventions
- Engaging a variety of users (eg, current A&F users, naive users, high performers, and low performers) to inform the development of A&F and its cointerventions
- Where resources are limited, focusing on high-value best-practice recommendations that influence engagement with the data (a necessary precursor to action and impact), including the following:
  - Addressing the credibility of the data
  - Including indicators that physicians value and perceive as actionable
  - Recommending actions consistent with the established goals and priorities

**Limitations**

First, the transferability of our findings is limited given the context, focused sample, and sampling approach (ie, 17 family physicians in Ontario who voluntarily signed up for the A&F report, including 10/17, 59% who were part of a family health team) as well as the specificity of the QI intervention examined. The way the A&F report was delivered in our study was a passive and solitary approach whereby physicians accessed their reports in confidence via email. A&F initiatives that support the creation of space for physicians to discuss the data with colleagues or a credible source and enable greater understanding and actions for improvement may be perceived as more usable. Methodologically, no member-checking process was undertaken to validate data interpretation among the research participants. However, we held peer debriefing meetings with the research team supported by senior researchers to review the data analysis and findings as well as discuss the interpretation of the findings. Finally, to address the change in research objective as mentioned previously, we described the research process in a transparent way and went back to the data to analyze and interpret them consistently to answer the research questions.

**Conclusions**

This study found that esthetic design changes played a minor role in how family physicians used the A&F report. The usability of A&F appears to depend on recipients’ perceptions of whether the quality indicators are important, accurately measured, and controllable through feasible clinical actions. Those who found the A&F report useful did so because they felt that it was aligned with the goals and priorities of their practice as a whole. Other family physicians might benefit from cointerventions to facilitate the integration of A&F into the workflow and build capacity to interpret the data and undertake practice-level actions accordingly. Health system administrators and clinicians should work together to optimize alignment between the report and the priorities of end users.
Acknowledgments

The authors would like to acknowledge and thank the family physicians who participated in this project. The authors would also like to acknowledge the staff of Health Quality Ontario for their help in making this work possible. They are grateful to Holly Witteman and David Flaherty, who both revised the first draft of this paper. Finally, the authors thank the research team members who were involved in previous data analysis cycles: Natasha Kithulegoda and Beth Bosiak. This study was funded by the Ontario Strategy for Patient-Oriented Research Support for People and Patient-Oriented Research and Trials Unit, which is supported by the Canadian Institutes of Health Research and the Province of Ontario. The funder was not involved in the design of the study or writing of this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Screenshot of the reports.

[PDF File (Adobe PDF File), 328 KB - humanfactors_v10i1e38736_app1.pdf]

Multimedia Appendix 2

Interview guide.

[PDF File (Adobe PDF File), 93 KB - humanfactors_v10i1e38736_app2.pdf]

Multimedia Appendix 3

Alignment between best practices, Clinical Performance Feedback Intervention Theory (CP-FIT) constructs, target of the redesigned report, and physicians’ engagement with the report.

[PDF File (Adobe PDF File), 215 KB - humanfactors_v10i1e38736_app3.pdf]

References


Abbreviations

A&F: audit and feedback
COREQ: Consolidated Criteria for Reporting Qualitative Research
CP-FIT: Clinical Performance Feedback Intervention Theory
QI: quality improvement
Patients’ Information Needs Related to a Monitoring Implant for Heart Failure: Co-designed Study Based on Affect Stories

Abstract

Background: RealWorld4Clinic is a European consortium that is currently developing an implantable monitoring device for acute heart failure prevention.

Objective: This study aimed to identify the main issues and information needs related to this new cardiac implant from the patients’ perspective.

Methods: A total of 3 patient collaborators were recruited to help us design the study. During 4 remotely held meetings (each lasting for 2 hours), we defined the main questions and hypotheses together. Next, 26 additional interviews were conducted remotely to test these hypotheses. During both phases, we used affect stories, which are life narratives focusing on affect and the relationship between patients and the care ecosystem, to highlight the main social issues that should be addressed by the research according to the patients.

Results: Context of diagnosis, age, and severity of illness strongly influence patient experience. However, these variables do not seem to influence the choice regarding being implanted, which relies mostly on the individual patient’s trust in their physicians. It seems that the major cause of anxiety for the patient is not the implant but the disease itself, although some people may initially be concerned over the idea of becoming a cyborg. Remote monitoring of cardiac implants should draw on existing remote disease management programs focusing on a long-term relationship between the patient and their medical team.

Conclusions: Co-design with affect stories is a useful method for quickly identifying the main social issues related to information about a new health technology.

Introduction

Background

Heart failure is a serious chronic condition and the leading cause of hospitalization in Europe for people aged >65 years. Because of the aging population, heart failure is considered a major threat to health care systems [1]. Several research projects have therefore been started in recent years aiming at improving heart failure care by using information and communication technologies [2]. In France alone, 6 telemedicine projects were undergoing development in 2018 [3-5], when the first national telemedicine program (ETAPES) was set up to develop and fund real-life remote monitoring [6].
The main issue regarding heart failure management is the prevention of cardiac decompensation, a sudden and life-threatening aggravation of the symptoms that is responsible for frequent hospitalizations of patients with this complication. For now, the detection of cardiac decompensation is mostly based on symptoms reported by the patients, notably weight gain. Physiologic signals recorded by implantable devices would allow earlier detection, resulting in lower rates of hospitalization [7].

**Objectives**

RealWorld4Clinic is a research consortium supported by European Institute of Innovation and Technology (EIT) Health that aims to develop MyHeartSentinel, an implantable connected device that could diagnose acute decompensated heart failure 30 days in advance, based on daily recordings of cardiorespiratory data [8]. A unique feature of RealWorld4Clinic is that it involves several researchers in humanities and social sciences, including the Ethics & AI Chair of the Multidisciplinary Institute in Artificial Intelligence of Grenoble Alpes University in Grenoble, France. The objective is to address the ethical, legal, and social issues raised by this new connected medical device early in the innovation process. This approach is inspired by works on ethical health technology assessment [9-11], with an emphasis on patient and public participation [12].

In this paper, we present research that aimed to identify patients’ information needs concerning MyHeartSentinel. For this purpose, we need to better understand patients’ perspectives on heart failure, implants, and remote monitoring.

**Methods**

**Overview**

Our study is divided into 2 parts. First, we co-designed the main research questions and hypotheses with a small team of patients who were interested not only in following the project but also in collaborating with researchers. Second, we strengthened these hypotheses via a qualitative study involving a wider panel of patients.

In both parts of the study we used affect stories, which are life narratives focusing on affect and relationships. Affect is a significant but long-overlooked part of human experience that is now receiving growing interest [13,14], notably in the design field [15]. By affect, we mean any affective phenomena, including feelings, moods, emotions, and attitudes [16]. These affective phenomena are central to social interactions and meaning-making processes [17]. Taking them into consideration is therefore very useful to analyze what matters to patients, what difficulties they face, and how to co-design with them [18].

**Part 1: Co-design of the Main Research Questions and Hypotheses**

To recruit our patient collaborators, we contacted RESIC38. RESIC38 is a health network dedicated to heart failure based at the Grenoble Alpes University Hospital. It is in charge of organizing patient pathways and maintaining a therapeutic patient education (TPE) program. It regularly organizes individual or group sessions on various topics, such as “My daily medications,” “Traveling comfortably,” and “Sexuality with a chronic condition.” TPE is an approach in the field of chronic condition management that promotes multidisciplinary and patient-centered care [19]. Since 2009, the French National Authority for Health (Haute Autorité de la Santé) has accredited TPE programs that follow its guidelines. The role of TPE is not only to inform patients but also to help them to adapt medical instructions to their daily lives. TPE therefore promotes patient empowerment and a paradigm shift in the relationship between patient and health care professional [20,21].

The director of RESIC38 conveyed our request to 3 active members of the network—3 men aged 56, 73, and 76 years—whom he considered capable of helping us in our research project. Starting in March 2021, several remote meetings were organized approximately once a month between the authors of this paper and these 3 patients. In the first session, each patient told us his story about heart failure. The patients were asked to expand on the various affective phenomena and social relationships they had experienced during their patient pathway and care pathway. This first session, which was combined with a literature search, allowed us to propose 4 main research questions and a set of hypotheses, which were refined during the second session (Multimedia Appendix 1). In the third session, we proposed a methodology to test these hypotheses, which consisted of collecting evidence from different sources: interviews with patients and health care professionals, literature searches, patient associations’ websites, and health forums. We also discussed the best ways to recruit new interviewees, patients, and health care professionals. At the fourth meeting, we presented and discussed the results of the first interviews. Each session was recorded, and we listened to the recordings to write the minutes, which were then sent to all participants.

**Part 2: Validation of the Co-designed Hypotheses via Qualitative Interviews**

Meanwhile, interviews were undertaken to test the co-designed hypotheses. These consisted of nondirected affect stories, completed with some questions. Our interview guide is presented in Multimedia Appendix 2.

Of the 26 interviewees, 19 (73%) were recruited from 2 health networks dedicated to heart failure: RESIC38 (n=8, 42%) and the cardiac unit of Hôpital Privé Le Bois in Lille, France (n=11, 58%). Both provide individualized and multidisciplinary follow-ups with their patients, including drug treatment optimization and patient education, but only RESIC38 organizes group education sessions as part of an official TPE program. Of the remaining 7 participants, 4 (57%) were members of patient associations, and 3 (43%) were contacted having been identified via their relevant posts on social networks.

The sample characteristics are summarized in Table 1. Of the 26 participants, 16 (62%) were men, and 10 (38%) were women. The average age was 65 (SD 17; range 21-89) years. Most (21/26, 81%) of the patients had already been implanted with a medical device with remote monitoring capabilities, and some (12/18, 67%) were part of a remote follow-up program using connected objects. Most (21/26, 81%) of the patients were living with heart failure, and 12% (3/26) had experienced it before...
receiving a heart transplant. The (2/26, 8%) exceptions were patients with suspected cardiac disease who had contacted us after receiving the recruitment advertisement by mistake. One of these interviews was particularly interesting for us because an implantable loop recorder was mentioned. This is a diagnostic tool used to record cardiac data, but it cannot deliver electrical impulses to regulate the heartbeat. It is close to the implant MyHeartSentinel in terms of form and implantation procedure.

Table 1. Participant characteristics (N=26).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
</tr>
<tr>
<td>Man</td>
<td>16 (62)</td>
</tr>
<tr>
<td>Woman</td>
<td>10 (38)</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
</tr>
<tr>
<td>&lt;50</td>
<td>4 (15)</td>
</tr>
<tr>
<td>50-65</td>
<td>8 (31)</td>
</tr>
<tr>
<td>&gt;65</td>
<td>14 (54)</td>
</tr>
<tr>
<td>Professional activity</td>
<td></td>
</tr>
<tr>
<td>Active</td>
<td>5 (19)</td>
</tr>
<tr>
<td>Unemployed</td>
<td>6 (23)</td>
</tr>
<tr>
<td>Retired</td>
<td>15 (58)</td>
</tr>
<tr>
<td>Socioprofessional group (before retirement or disablement)</td>
<td></td>
</tr>
<tr>
<td>Farmer</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Artisan, merchant, business executive</td>
<td>3 (12)</td>
</tr>
<tr>
<td>Upper managerial or intellectual occupation</td>
<td>7 (27)</td>
</tr>
<tr>
<td>Intermediate profession</td>
<td>8 (31)</td>
</tr>
<tr>
<td>Employee</td>
<td>5 (19)</td>
</tr>
<tr>
<td>Blue-collar worker</td>
<td>3 (12)</td>
</tr>
<tr>
<td>Time (years) since first cardiac follow-up</td>
<td></td>
</tr>
<tr>
<td>&lt;2</td>
<td>5 (19)</td>
</tr>
<tr>
<td>2-10</td>
<td>7 (27)</td>
</tr>
<tr>
<td>10-30</td>
<td>11 (42)</td>
</tr>
<tr>
<td>&gt;30</td>
<td>3 (12)</td>
</tr>
<tr>
<td>Implants</td>
<td></td>
</tr>
<tr>
<td>Defibrillator</td>
<td>12 (46)</td>
</tr>
<tr>
<td>Pacemaker</td>
<td>5 (19)</td>
</tr>
<tr>
<td>Heart transplant</td>
<td>3 (12)</td>
</tr>
<tr>
<td>Implantable loop recorder</td>
<td>1 (4)</td>
</tr>
<tr>
<td>None</td>
<td>5 (19)</td>
</tr>
</tbody>
</table>

There was no prior relationship with any participant. An informed consent document was sent to each participant and re-explained at the beginning of each interview. All of the participants agreed to be recorded. They were interviewed remotely for approximately 1 hour, either by videoconferencing or by telephone. On 3 occasions the participant’s partner was also present and intervened during the interview.

Each interview was replayed once and summarized by the interviewer (AD). The portions referring to affective experience or relationships were transcribed verbatim.

Information that could be used to identify the patients was either generalized (eg, city names were replaced by brief sociodemographic information) or anonymized (eg, in the case of physicians’ names).

Ethical Considerations
According to French legislation, our study did not require ethics approval because our aim was not to develop biological or medical knowledge. However, we sought and received approval from the multidisciplinary ethics committee of Grenoble Alpes University (CERGA-Avis-2021-24), which checked the
compliance of our interviews with the European General Data Protection Regulation.

Results

Overview

Our analysis is mostly deductive, based on the research questions and hypotheses codefined with the patient collaborators. In the following paragraphs, we present our results according to this reading grid (the hypotheses are presented in Multimedia Appendix 1). We have selected some representative quotes, which have been translated from French into English.

Question 1: Are There Patient Profiles for Which the Implant Is (or Is Not) Appropriate? In Particular, Is It Necessary for the Patient to Accept the Disease Before Entering a Monitoring Program?

Overview

The experience of living with heart failure seemed to vary significantly from one interviewee to another. To better understand these different perspectives, we analyzed similarities and dissimilarities among the affect stories. We identified 3 key factors that strongly influence the patient experience: context of diagnosis, age, and illness severity.

Context of Diagnosis

Some people became patients living with heart failure overnight after an emergency hospitalization for myocardial infarction or acute pulmonary edema. Thus, they discovered intensive care and the world of cardiology for the first time. On their return home, they had to learn to live with a new chronic condition, as observed by a patient:

I had never been sick. I mean severely ill. This was my greatest wealth...When they told me that I had to be anesthetized, I was terrified. I had never been anesthetized in my life!...All my life, I had never taken my blood pressure. So uh...I learned to do all that...I have the greatest difficulty with accepting myself as “being sick.” [P17, woman aged 80 years hospitalized a few months previously for myocardial infarction]

Others slide progressively into heart failure after years of cardiac follow-up. Their perception of the disease and the relationship with the medical community may therefore be quite different from those patients who take ill suddenly. A patient stated as follows:

I have seen people at the RESIC, who were on top form and suddenly...a shock. Before that, they could run, etc...I, however, have never been able to run the 100 meters, you see? So it didn’t change my life. [P9, man aged 86 years diagnosed with a heart defect in childhood]

Between these 2 extreme examples, there is a great diversity of trajectories. In particular, many (8/26, 31%) of the patients knew that they had a family history of cardiac issues, which nevertheless did not prevent them from being startled by their first hospitalization. This family history is a source of concern, but it is also a means by which they can picture themselves in the future, as explained by a patient:

When the cardiologist told me that the results of my ultrasound were not good, I collapsed, because all the images of my ill father were brought back and I thought: “This time, it’s my turn.”...When I understood I had the same pathology (we compared the medical records), I knew that I would not escape a heart transplant. And thus, I had the time to prepare mentally, while my father didn’t. He was so afraid that he gave up. He gave up and died very quickly. [P7, man aged 45 years and heart transplant recipient]

Age

Although heart failure is very frequent among those aged >65 years, it can occur at any age [22]. We observed significant differences between the experiences of younger patients and those of older patients.

Professional activity is a major issue for younger patients. They often need accommodation at work or professional retraining, especially if they have a physical job. Sometimes they are not able to work any longer and instead have to survive on disability allowance. Another concern is parenthood because pregnancy is discouraged for patients with a cardiac condition, and taking care of children is more difficult because of the physical limitations and uncertainty associated with the disease. For these patients, heart failure is an invisible disability that is hard to reconcile with social conventions. Sometimes, they are reluctant to use the assistance to which they are entitled, such as reserved parking places, because of what people might say. Therefore, they must learn to deal both with the need not to look ill and the need to conserve their limited energy.

Fatigue and breathlessness are more easily accepted among older adult patients or even downplayed. As in the case of younger patients, they feel the loss of their physical capabilities, but they do not attribute it only to heart failure. They often experience several pathologies: not only cardiovascular issues (such as hypertension) or diabetes but also respiratory illness, sleep apnea, visual or auditory impairment, osteoporosis, loss of balance, dementia, and so on. As a consequence, they tend to take many medications, which increases the risks of unwanted side effects and unobservance. Optimizing their treatment requires many trade-offs. To take just 1 typical case: a participant (a man aged 83 years) explained that he was advised by his cardiologist to stop diuretics to preserve renal function, but very soon he had to resume his usual medication because of water retention.

Heart failure is not always the patient’s main concern, especially if it is at an early stage; for example, a patient aged 73 years, who had been successfully treated for cancer 10 years previously, reported that when receiving his blood test results, he was more worried about tumor markers than about heart failure markers. Sometimes, the patients are also caregivers for their partner, which causes them considerable anxiety and affects their finances if their partner needs to be moved to a nursing home.
Illness Severity

Heart failure severity is usually evaluated according to the New York Heart Association functional classification system [23]. This system consists of 4 classes based on the symptoms reported by patients and how these symptoms affect their daily lives by limiting their physical activities. A patient in class I does not show any symptom of cardiac impairment, whereas a patient in class IV is unable to undertake any physical activity (including walking) and may experience fatigue, palpitation, dyspnea, or angina pain even at rest. Disease management aims at reducing these symptoms and slowing down disease progression. If the disease is advanced and resistant to treatment, heart transplantation is the last resort, but it is a rare and dangerous operation, reserved for patients with the greater benefit-risk ratio.

It is worth noting that our study participants did not mention their New York Heart Association class. However, they frequently talked about their left ventricular ejection fraction (LVEF) value and how it had evolved since their diagnosis. LVEF describes the efficacy of the heart in pumping blood. The LVEF value may rise with disease management or fall in cases of aggravation.

The aforementioned factors influence the difficulties faced by patients and thus their individual level of illness acceptance, that is, their psychological adaptation to the illness [24]. It is clear from the interviews that this acceptance takes time and that it is not always possible for the patients to accept their illness. Indeed, some of them cannot bear the thought of losing their physical capabilities; for example, an interviewee told us that before his transplantation he had tried to keep cycling as though he was not ill at the risk of aggravating his heart condition. Another explained that he was working part time as a consultant. Should he stop working, he stated, it would be “the end of everything.” However, neither patient was reluctant to participate in management of his disease or to test new treatments; on the contrary, they were keen to do so to improve their physical condition. This suggests that illness acceptance is not essential for a patient to accept a monitoring implant. On the contrary, it could be seen at first glance as a way to escape the illness by delegating self-monitoring to the device. This would not necessarily be a problem, provided that remote monitoring works and that the patients have access to a TPE program when they are ready to be more involved in the management of their disease. In this case, the implant would be akin to Ariadne’s thread, connecting the patients to their health care teams and maybe even to their peers.

Indeed, most of the interviewees thought that interaction with other patients was very important. These interactions could either be supervised by a medical team as part of a TPE program or initiated by the patients themselves. We used a thematic analysis to understand what they were seeking in these interactions (Multimedia Appendix 3).

Question 2: What Are the Determining Factors That Would Lead Someone to Accept or Reject a Monitoring Implant?

At the time of the interviews, most (18/21, 86%) of the patients who had experienced the implantation of a cardiac prosthesis (including the 3 persons who wore a defibrillator before receiving a heart transplant) seemed to consider it as just another step on their patient pathway. They mentioned it briefly and sometimes did not even do so until they were questioned about their follow-up. When they expressed their feelings, they thought first about what this implant meant regarding their health condition:

[About her defibrillator] That’s what has marked my transition to serious heart problems. [P4, woman aged 50 years and heart transplant recipient]

[After my first hospitalization], it was a second shock, more violent, because I thought: “It’s not a pacemaker, because apparently my heart is beating, but it might race. [P7, man aged 45 years and heart transplant recipient]

[About her pacemaker] It is all the better for me because it means that they think that my health is good enough to benefit from it. [P23, woman aged 83 years]

These patients consented quickly to the implantation as part of their treatment, trusting their cardiologist’s advice. Most (15/21, 71%) of them had been implanted with a defibrillator and mentioned that the device gave them a sense of safety. For those who had had to wear a cardiac LifeVest for months, the implantation was even a relief because they no longer had to live with a wearable defibrillator day and night, as explained by a patient:

This LifeVest...It was horrible. It weighs two and a half kilos, and you always need to carry it. When I was walking I carried it...It is far better to insert the defibrillator as I have it now. Because even at night I should keep it and sleep with it. It was not a panacea. [P21, man aged 75 years]

Only 14% (3/21) of the patients delayed their implantation for as long as possible: P10, P13 and P14 (see next section). They emphasized the importance of receiving moral support to overcome their concerns:

The only person who helped me, it was when I got my defibrillator: the Social Security and the physicians gave me a psychologist for three sessions...My cardiologist had told me a while ago that I should get a defibrillator. He told me that for three or four years: defibrillator, defibrillator...He is a super guy, so I said: “Well, we’ll see...” Finally, he gave me this defibrillator. Hum...it went well, but I took a moral blow anyway. I suddenly became much grayer. [P10, man aged 79 years]

At 50 years old I was not very happy about having a foreign object in my body...I was wondering if my physical abilities would degrade in relation to this implant. It’s very important to have moral support. I
did not have it. I insisted for three years on not being implanted and finally I accepted, and I am happy to have done it because I had a heart attack one year later. [P13, man aged 55 years]

This confirms our first hypothesis: the determining factor in the acceptance or refusal of an implant is the trusting relationship with a health care team. Oudshoorn [25] has even suggested that patients do not really have a choice because these implants are the present standard of care.

To go further, we analyzed the conditions required for this trusting relationship, based on both the positive and negative experiences reported by the participants. The emerging themes and subthemes are summarized in Multimedia Appendix 4, along with some quotations. Many of these can be linked to TPE, as suggested in our second and third hypotheses (Multimedia Appendix 1).

**Question 3: What Are the Main Sources of Anxiety Related to the Implantation of a Monitoring Device?**

It is already known that many cardiac implant wearers face anxiety or depression [26,27]. However, as previously mentioned, our study participants did not talk very much about their implants; instead, their feelings of anxiety seemed to be linked more to the severity and unpredictability of their illness. In other words, they are afraid of dying:

- *“I didn’t dare to do sport too much on my own anymore because I was really scared of an accident, of my heart racing, because I had been told it was the main concern.” [P7, man aged 45 years and heart transplant recipient]*
- *“The longest thing was...I was back home and whenever I fell asleep at night I thought: “Will I wake up tomorrow?” Because it was a close one. The odds were against me and I was lucky to survive...So, afterwards, you stay a long time with this idea: “Will I wake up tomorrow?” While thinking: “Anyway, what can I do about it?” [P25, man aged 73 years]*
- *“I am old, but not in my head. And that’s what I struggle to accept, because I love to tinker and things like that...Morally, it is like a blow because I am always wondering: “How will it evolve? Can I plan something in three months...six months...?” I don’t have the answer.” [P21, man aged 75 years]*

The sources of anxiety listed in our hypotheses were mentioned but as inconveniences rather than as deterrents, which is consistent with the results of prior studies based on the Florida Patient Acceptance Survey [28-30]. Our interpretation is that the patients’ concerns are mostly linked to their illness and its consequences for their daily lives and mortality, rather than to the implant itself. However, our interviewees have either never experienced defibrillator shocks or experienced them only on very rare and appropriate occasions. The situation is certainly very different for patients experiencing multiple shocks [31,32].

Whether these results are easily transposable to monitoring implants is unclear. Implanting a monitoring implant under the skin is safer than implanting a pacemaker or a defibrillator, the leads for which can be a source of medical complications [33,34]. However, there is greater public awareness about pacemakers and defibrillators than about monitoring devices. Moreover, pacemakers and defibrillators actively contribute to the health of their wearer, whereas a cardiac monitor may seem excessive if it is presented only as a diagnostic tool; for example, a short paper in 2012 reported that among 1093 patients with kidney failure screened for a pilot study, 372 were found to be suitable, and only 8 were accepted to receive an implantable cardiac monitor [35]. Later studies were more successful, probably thanks to the miniaturization of the device [36]. Some research has even explored how the implantable loop recorder is perceived by patients [37,38].

To better understand patients’ motives for rejecting implants, another theme should be explored: the “foreign body” or cyborg theme. Indeed, some patients are concerned not only about the impact of the implant on their lives but also about its mere presence inside their body. Among our 21 study participants who experienced implanted medical devices, 2 (10%) had asked their cardiologists whether their implant could eventually be removed if they got better, or after their death. Another noted that his friends make fun of him by calling him a robot and that his defibrillator’s wires are visible on his x-ray images. In extreme cases, patients may perceive the implantation to be a dangerous operation and a threat to their human identity. This was the reaction of the interviewee P14 who wore an implantable loop recorder. This person was a retired nurse aged 70 years. Because she experienced transient ischemic attacks (ministrokes), she was sent by her referring physician to a cardiac rhythmologist to find the cause of these events. But the consultation went wrong: the patient was flabbergasted at the sight of the device and strongly disagreed with having this “foreign body” inside her. She blamed the cardiac rhythmologist for running out of patience when she started asking questions, as if everything were already decided. After her referring physician insisted, she finally agreed to meet another cardiac rhythmologist and ultimately consented to the implantation.

**Question 4: What Is the Impact of a Monitoring Device on the Patient Pathway?**

Different experiences of remote follow-up were reported during the interviews.

The patients (n=11) recruited via the Hôpital Privé Le Bois were (or had previously been) part of a remote monitoring program. Each day they measured their weight and blood pressure and filled out a symptom survey via a set of connected objects provided by the hospital. Even those who had limited experience of IT had no difficulty using these devices. As part of a home return assistance program (PRADO, the French acronym for Posthospitalization Home Return Assistance Program), some (2/11, 18%) of them also received visits from a nurse during the 2- to 6-month period after their last hospitalization. These participants were genuinely surprised at the quality of their care. One person was a retired nurse aged 70 years. Because she experienced transient ischemic attacks (ministrokes), she was sent by her referring physician to a cardiac rhythmologist to find the cause of these events. But the consultation went wrong: the patient was flabbergasted at the sight of the device and strongly disagreed with having this “foreign body” inside her. She blamed the cardiac rhythmologist for running out of patience when she started asking questions, as if everything were already decided. After her referring physician insisted, she finally agreed to meet another cardiac rhythmologist and ultimately consented to the implantation.

The patients (n=11) recruited via the Hôpital Privé Le Bois were (or had previously been) part of a remote monitoring program. Each day they measured their weight and blood pressure and filled out a symptom survey via a set of connected objects provided by the hospital. Even those who had limited experience of IT had no difficulty using these devices. As part of a home return assistance program (PRADO, the French acronym for Posthospitalization Home Return Assistance Program), some (2/11, 18%) of them also received visits from a nurse during the 2- to 6-month period after their last hospitalization. These participants were genuinely surprised at the quality of their care. One person was a retired nurse aged 70 years. Because she experienced transient ischemic attacks (ministrokes), she was sent by her referring physician to a cardiac rhythmologist to find the cause of these events. But the consultation went wrong: the patient was flabbergasted at the sight of the device and strongly disagreed with having this “foreign body” inside her. She blamed the cardiac rhythmologist for running out of patience when she started asking questions, as if everything were already decided. After her referring physician insisted, she finally agreed to meet another cardiac rhythmologist and ultimately consented to the implantation.

https://humanfactors.jmir.org/2023/1/e38096

JMIHumFactors 2023 | vol. 10 | e38096 | p.446

(page number not for citation purposes)
Every morning, I take my blood pressure, I weigh myself, and I send all that to the monitoring center. I don’t know where it is, I don’t know who...who takes my stuff. But it works quite well, because on occasion I left for a weekend or a couple of days and I didn’t bring all these things which were a little heavy in the suitcase with me. And they didn’t miss! The nurse called me saying: “Mister X, are you feeling well?” So it’s well followed up. [P18, man aged 75 years]

Every morning, around 11 AM, 11:30 AM, they receive all the results. If necessary, they call me. If it’s not necessary, well, they don’t call. And if there is nothing at all during one or two weeks, they call me anyway. It’s more to catch up. They said: “Don’t worry, there is nothing, it’s just to check on how you feel whether everything alright, whether you’re not anxious.” This phone call is really...a comforting touch. [P24, man aged 61 years]

The patients thus feel reassured by this program. Of the 11 patients, 3 (27%) mentioned that they were more motivated to watch their weight because someone was watching over them. They stated that they can even feel empowered by this approach:

In the clinic, I was in a bit of a strange psychological state. Because every morning, I faced myself in the mirror as a patient. And when I received visits from these ladies, I was like, I am with you right now: more like making a contribution to an action, on something. You see? Well, it was a great help anyways...It’s wonderful because we feel surrounded and supported. You see, it’s like a kind of...partnership. I live it as a kind of partnership. Everyone has a place, of course, I am not a cardiologist or a specialized nurse. But it’s a...a dialogue actually. It allows people to be an actor of their health, we can say it like that. [P17, woman aged 80 years]

Of the 11 patients in this program, 10 (91%) also had a pacemaker or a defibrillator whose proper functioning was monitored by the same nurses. Opinion on remote monitoring seemed to be more divided among the other patients with implants outside of this program (n=12). Of these 12 patients, 2 (17%) had actually been contacted by the hospital because of a malfunction of their implants and estimated that they were well monitored, but 5 (42%) expressed doubts that the monitoring was really effective because they had never received a call when they experienced arrhythmia or even a shock from their defibrillator. They stated that they regretted having to call the hospital themselves to obtain information:

I think it could be a useful tool if it was monitored. Sometimes I had alerts, but no phone call. Whereas I was told, “As soon as we’ll see an episode, we’ll inform you.” I needed that to be reassured somehow, but what seemed odd to me was that when there were alerts, I was the one who had to seek information, instead of information coming to me thanks to the monitoring people. [P7, man aged 45 years and heart transplant recipient]

As has already been shown by Skov et al [39], many patients are not satisfied with a “No news is good news” approach. They need to directly and repeatedly experience that there is actually “someone at the end of the wire” to trust the remote monitoring [40]. Moreover, in the absence of a program dedicated to remote monitoring, the follow-up of cardiac implants seems to be less diligent and coordinated because it is provided by nurses who have other duties to attend to. A Swedish research study on remote monitoring showed that nurses struggle to manage alerts from multiple interfaces (one for each manufacturer) and that the time required to do so was not always acknowledged by their managers [41].

Another difference between the 2 patient groups could be seen in their access to their health data. The patients from the Hôpital Privé Le Bois have access to their measures and their history, which became a conversation topic with the health care professionals and a learning opportunity, as noted by the interviewees:

Suddenly my blood pressure rose from 9/8 to 12/8. I was wondering: “What does it mean?” It was strange...I talked about it with the nurse and we reviewed the previous records together and it was alright...I realized that maybe I should be more careful about what I do between my breakfast and my measurement. [P17, woman aged 80 years]

I saw the nephrologist yesterday and he said I was taking my blood pressure too fast and I should be more relaxed. Surely my blood pressure was lower this morning. But I may have made a mistake because I took my medications first...I think I should take my blood pressure before. [P23, woman aged 83 years]

By contrast, patients who wear a cardiac implant have no free and immediate access to their own data, the analysis of which is performed by specialized cardiologists whom they meet only once or twice a year. Even if the patients write down dates and times when they feel an abnormal sensation, these notes do not match the physician’s observations. Conversely, abnormal recordings are very difficult to link with their experience, as noted by a participant:

My defibrillator records tachycardia episodes. Surprisingly, it does not correspond to a particular fatigue or overactivity...Sometimes I feel bad. I can’t explain what it is, but I don’t feel well. So I note it and when I go to the control visit, I ask, “What happened at this time?” “Nothing. Everything was fine.” But two months later, bang! There is a burst. [P3, man aged 76 years]

Eventually, the patients need to use additional measuring instruments such as smartwatches to monitor their heart rate when they exercise and avoid experiencing a shock from their defibrillator. In the case of an audible alert from their device, they may not recognize it immediately or know what to do, as stated by a participant:

One day, it’s rare but it happens, the wire to the heart broke. You see? Of course, I didn’t know it, but I felt bad. It was the first time I was hearing a small ring...
inside me. I told my wife: “Do you hear a ring? It’s like a phone ringing.” And she said: “No, no. There is no phone ringing.” But it was inside, you see. It’s weird when you’re not used to it...I went to the shower. I grabbed the shower head, or something else. I passed out and I woke up sitting in the shower. That’s when I understood...While my daughter was driving me to the hospital, I collapsed maybe five or six times because I didn’t know that each time I raised my left arm, the contact was lost. [P8, man aged 63 years]

It seems that TPE is completely overlooked in the design of current cardiac implants, which focus solely on sustaining heart function. The development of a new monitoring device could provide an opportunity for patient empowerment. However, Lomborg et al [32] have shown that access to self-tracking data has ambivalent effects. It may be a tool to promote self-care, but it can also be a cause of frustration and distress when the patients are not able to associate the data with their sensations and emotions. This issue could be even more acute with implantable medical devices [42], whose measures are supposed to be accurate and reliable. It is also expected that, in some cases, the data will have very sensitive implications for the patient. If their illness is worsening, they should be informed first by a physician, not by an application.

**Discussion**

In this final section, we will discuss not only the implications of this study on the development of an information medium related to MyHeartSentinel but also the transformation in patient monitoring brought about by the development of such health implants.

**Pros and Cons of an Implantable Monitoring Device**

MyHeartSentinel is an implantable device, which has both advantages and drawbacks. Among its advantages, it allows trustworthy and automatic measurements. Compared with the CardioMEMS implant [43,44], this subcutaneous implant is less invasive, and once the gateway is installed, data transmission will not require any commitment (a priori) on the part of the patient. It is already known that adherence to self-care is an important issue in the field of chronic conditions [45], including heart failure [46-48]. Therefore, it seems interesting to use low-invasive implants to deliver medication [49] or, in our case, to monitor patients living with heart failure. Ideally, a team of health care professionals should be dedicated to remote monitoring to quickly respond to any alert and to support their patients in disease management.

In terms of drawbacks, we can expect patients to be reluctant to agree to the implantation. Diverse reasons have already been mentioned in our paper: fear of surgery, threat to personal privacy, transformations in daily life and the relationship with the health care team, and fear of becoming a cyborg. The success of this particular implant will depend on whether SentinHealth (the medical technology start-up developing MyHeartSentinel) will be able to convince cardiologists that its device is relevant, both in terms of medical outcomes and organizational routines.

**Patient Empowerment and Co-design as a Condition to Effective Disease Management**

Close follow-up of medical data is certainly useful, but it will not be sufficient to improve heart failure management. Our interviews as well as reviews on remote monitoring showed that the success of remote monitoring is highly context dependent [2-5,50-52]. The important thing is not just which remote monitoring system should be used but also how it should be used and by whom. It seems essential to co-design not only with patients but also with every stakeholder as much as possible, and the process should include the technical device itself as well as the clinical routines and information supports [53-57].

Another challenge, identified by Greenhalgh et al [2], is the possibility of “tinkering” with remote monitoring to adjust to each particular situation, which may seem contradictory given the values of standardization and quantitative performance that are generally associated with automated systems. In our study, we collected a great variety of narratives, which we analyzed through the prism of 3 key variables: context of diagnosis, age, and illness severity. Analysis through other lenses such as gender, social class, psychology, culture, or isolation would certainly lead to interesting conclusions [58-61]. In the face of this complexity, it seems difficult to define relevant patient profiles. Rather than a one-size-fits-all approach focused on one of these profiles, we think that it is better to consider a modular and customizable follow-up program; for instance, patients would decide with their physicians which information and services they should have access to. This could be renegotiated over time, according to patients’ readiness to participate in the remote monitoring program. In this case, clinical evaluation should not be limited only to hospital admission and survival rates but should also include patient-centered outcomes, self-defined by each patient [52].

**Limitations**

Although diversified by gender, sociocultural background, location, age, and time since diagnosis, our sample is not representative of the entire population of patients living with heart failure. This is a methodological problem observed in many studies requiring patient involvement [55,56,62]. The 3 patient collaborators as well as most of the interviewed patients are highly educated (17/26, 65%), able to speak with ease (23/26, 88%), and are willing to share their experiences. Issues related to health care access or literacy may therefore be underestimated.

Moreover, our recruitment method did not allow us to access patients who are uninformed or in denial about their disease, which is certainly a big issue in heart failure. According to the ICPs2 survey administered in 2018 to nearly 800 patients hospitalized for acute heart failure at 40 centers, 1 in 3 patients was not able to name the disease [63]. It should be noted that MyHeartSentinel was still in development at the time of publication of this paper. We based our study on interviews with patients who had had experiences of heart failure, implants, and remote monitoring, but none of them wore the new implant yet. This study was therefore undertaken to...
understand what kind of conditions would be required for patients to accept the new implant, and the results will need to be confirmed with future patients.

**Conclusions**

We have presented research aimed at identifying the issues and information needs related to an implantable monitoring device for patients with heart failure. After co-designing the hypotheses of the study with a small team of patient collaborators with a methodology based on affect stories, we tested the hypotheses via 26 additional interviews. Most of the initial hypotheses were validated, and some were rephrased or completed by our observations. None was discarded. This confirms that co-design with affect stories is an effective method for quickly identifying social issues related to a new health technology.

We found that the monitoring implant should be conceived primarily as a mediation instrument, rather than as a quantified self tool, that facilitates illness acceptance and communication between patients and health care professionals. The results of this study will be used to design the prototypes of an information module in collaboration with user experience designers at SentinHealth.

**Acknowledgments**

This publication is part of RealWorld4Clinic, an activity that has received funding from EIT Health. EIT Health is supported by the European Institute of Innovation and Technology, a body of the European Union. The authors would like to thank each contributor to RealWorld4Clinic’s education and training package: Daniel Pagonis, Philippe Cinquin, Thierry Ménissier, and Charles-Clemens Rüling. Special thanks to Annabelle Drault and Raphaël Koster, who shared their expertise when we were defining our research questions and hypotheses. This study would not have been possible without the help of Baptiste Barjhoux and Murielle Salvat from RESIC38, as well as Jessica Dayez and Frédéric Mouquet from the Hôpital Privé Le Bois. The authors are also very grateful to all study participants, especially our patient collaborators.

**Conflicts of Interest**

None declared.

---

**Multimedia Appendix 1**
Research questions and hypotheses co-designed with the patients.

[DOCX File, 16 KB - humanfactors_v10i1e38096_app1.docx]

**Multimedia Appendix 2**
Interview guide.

[DOCX File, 18 KB - humanfactors_v10i1e38096_app2.docx]

**Multimedia Appendix 3**
Motives for interacting with other patients.

[DOCX File, 17 KB - humanfactors_v10i1e38096_app3.docx]

**Multimedia Appendix 4**
Health care professionals' qualities.

[DOCX File, 22 KB - humanfactors_v10i1e38096_app4.docx]

**References**


21. Davat & Martin-JuchatJMIR HUMAN FACTORS


43. Pour-Ghazi I, Hana D, Raja J, Ibebuogu UN, Khouzam RN. CardioMEMS: where we are and where can we go? Sensors (Basel) 2021 Jan 28;21(3):887. [Full text] [doi: 10.3390/s21030887] [Medline: 33525556]


Abbreviations

EIT: European Institute of Innovation and Technology

LVEF: left ventricular ejection fraction

PRADO: Posthospitalization Home Return Assistance Program

TPE: therapeutic patient education

©Ambre Davat, Fabienne Martin-Juchat. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 23.01.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
The Acceptability of Technology-Based Physical Activity Interventions in Postbariatric Surgery Women: Insights From Qualitative Analysis Using the Unified Theory of Acceptance and Use of Technology 2 Model

Pierre Thérouanne1*, PhD; Meggy Hayotte2*, PhD; Florent Halgand1, MSc; Fabienne d’Arripe-Longueville2, PhD

1Université Côte d’Azur, LAPCOS, Nice, France
2Université Côte d’Azur, LAMHESS, Nice, France
*these authors contributed equally

Corresponding Author:
Pierre Thérouanne, PhD
Université Côte d’Azur, LAPCOS
Laboratoire d’Anthropologie et de Psychologie Cliniques, Cognitives et Sociales Campus Saint Jean d’Angely / SJA3 / MSHS Sud-Est
25 avenue François Mitterrand
Nice, 06300
France
Phone: 33 4 89 15 23 49
Email: pierre.therouanne@univ-cotedazur.fr

Abstract

Background: Bariatric surgery offers an opportunity for physical activity (PA) promotion due to patients’ increased ability to engage in PA. Technology-based PA interventions are promising tools for promoting PA to support patients in this key period. The Unified Theory of Acceptance and Use of Technology 2 (UTAUT2) model is a recognized theoretical model for examining technology acceptability. Although a previous study reported that 92% of women with obesity have high acceptability of at least one technology-based PA intervention, little is known about the factors that lead to different levels of acceptability between technologies and therefore the reasons for choosing a preferred intervention.

Objective: The purpose of this study was to (1) characterize the acceptability of 3 technology-based PA interventions (ie, telehealth, active video game, mobile app) in the context of bariatric surgery, and (2) explore patients’ preference motives. This study, using a qualitative design, examined the suitability of the UTAUT2 model in this specific context.

Methods: Participants (n=26) read written French descriptions of the technology-based PA interventions with illustrations and chose their preferred intervention. Semidirective interviews were conducted to explore the reasons for their choice of the preferred intervention, notably using the UTAUT2 framework. Data were analyzed based on inductive and deductive approaches.

Results: All participants who preferred a technology-based PA intervention (ie, active video game, n=10; mobile app, n=10; telehealth, n=6) expressed a behavioral intention to use it. In addition, some of them expressed a high behavioral intention to use another technology (ie, active video game, n=4; mobile app, n=1; telehealth, n=7). All the constructs of the UTAUT2 emerged during the qualitative interviews and were specified through subcategories. Additional constructs also emerged, especially other motivational factors.

Conclusions: This study showed that, in the context of technology-based PA interventions for postbariatric patients, the UTAUT2 is suitable, although additional motivational factors (which were not considered by the UTAUT2 model) should be considered.

(JMIR Hum Factors 2023;10:e42178) doi:10.2196/42178

KEYWORDS

acceptability; health technology; physical activity; obesity; UTAUT2
Technology-based physical activity (PA) interventions have been increasingly investigated in recent years to promote PA for vulnerable populations. These interventions have been used effectively to promote PA in the context of obesity care [1-4]. A recent meta-analysis confirmed that they were able to increase moderate-to-vigorous PA for women with obesity by approximately 25 minutes per week [5]. We also note an emerging interest in technology-based PA interventions in the context of bariatric surgery [6]. Bariatric surgery induces major weight loss that is perceived by patients as an increase in their ability to engage in PA [7]. In addition, PA is a favorable factor for long-term weight loss maintenance [8]. However, many postbariatric patients do not increase their PA and some even decrease it [9]. Women, especially young women, represent a higher proportion of bariatric surgery patients than men [10] and seem to be more prone to physical inactivity and sedentary behavior [11,12]. Thus, young women after bariatric surgery offer a good example of a chronic disease population in a key period to induce behavior change.

To this end, 3 categories of technology-based PA interventions can be recommended to patients: mobile technology (eg, mobile apps, wearable devices), game-based interventions (eg, active video games, exergames, serious games, augmented and virtual reality games), and computer- and internet-based interventions (eg, telehealth, email, websites, social media) [13-16]. Some technology-based PA interventions are more preferred (ie, more accepted) than others [17]. However, little is known about the preference motives of postbariatric surgery patients. Thus, characterizing the acceptability of technology-based PA interventions in this context would encourage the individualization of the recommendations for a given intervention based on the patient profile. Doing so would also provide engineers with information on patients’ preference motives that could guide them in adapting or developing new adapted technology–based PA interventions tailored to postbariatric surgery patients.

The reasons why some tools are chosen, accepted, and used more than others can be explained by models of acceptability [18]. Among the models, the Unified Theory of Acceptance and Use of Technology 2 (UTAUT2) [19] is today the most comprehensive, parsimonious, and powerful predictive model of the behavioral intention to use technology [20,21]. The UTAUT2 model is an extension of the UTAUT to a consumer context [19]. The model assumes that performance expectancy, effort expectancy, social influence, and facilitating conditions are key constructs that influence behavioral intention to use a technology or technology use [22]. The UTAUT2 incorporates 3 additional constructs, namely, hedonic motivation, price value, and habit [19]. The UTAUT2 constructs are defined as follows: (1) performance expectancy refers to “the degree to which using a technology will provide benefits to consumers in performing certain activities,” (2) effort expectancy refers to “the degree of ease associated with consumers’ use of technology,” (3) social influence refers to “the extent to which consumers perceive that important others (eg, family and friends) believe they should use a particular technology,” (4) facilitating conditions refers to “consumers’ perceptions of the resources and support available to perform a behavior,” (5) hedonic motivation refers to “the fun or pleasure derived from using a technology,” (6) price value refers to “consumers’ cognitive trade-off between the perceived benefits of the technology and the monetary cost of using it,” and (7) habit refers to “the extent to which an individual believes the behavior to be automatic” [19].

This model has been adapted into French in the context of eHealth [23]. Moreover, studies have recently investigated the relevance of the UTAUT2 model in certain chronic diseases. For example, in the case of diabetes, all UTAUT2 constructs were found to be relevant and 2 additional constructs, trust and perceived disease threat, also emerged as predictors of mHealth acceptability [24]. Several studies have also extended the UTAUT2 model to a variety of contexts that can be grouped into 6 categories: (1) different types of users, (2) different types of organization, (3) different types of technology, (4) different task types, (5) different times, and (6) different locations [25]. However, this model has rarely been used in the specific context of PA interventions, and even less so after bariatric surgery, which is a good example of a critical period for behavior change.

In the context of obesity care, including care for bariatric surgery patients, a latent profile analysis identified 2 acceptability profiles: (1) a high acceptability profile (ie, n=230 for telehealth, n=235 for active video game, and n=257 for mobile app), and (2) a low acceptability profile (ie, n=82 for telehealth, n=77 for active video game, and n=55 for mobile app) [17]. This study also demonstrated that these acceptability profiles were related to motivational factors (which were not considered by the UTAUT2 model). Although 92% of the women with obesity were in a high acceptability profile for at least one of the three technology-based PA interventions, this study did not account for the factors that led to different levels of acceptability between technologies. Therefore, it provided no information about the specificities of the different UTAUT2 constructs in the context of technology-based PA interventions for postbariatric surgery patients (ie, the items measuring the UTAUT2 constructs are generic and therefore not specifically tailored to this context), nor about their preference motives.

This study aimed to (1) characterize the acceptability of 3 technology-based PA interventions (ie, telehealth, active video game, mobile app) in the context of bariatric surgery; and (2) explore patients’ preference motives. Using a qualitative design, the study examined the suitability of the UTAUT2 model in this specific context.

**Methods**

**Procedure**

Individuals were invited to participate in this study in the waiting rooms for their routine postbariatric surgical care appointments in the South of France after participation in a previous quantitative study [17]. Eligible participants had read the written French descriptions of 3 technology-based PA interventions with illustrations in a counterbalanced order following a Latin-square design: active video game, mobile app, and telehealth (Multimedia Appendix 1). After reading the
descriptions, they classified the technology-based PA interventions according to their preferences and were asked if they would be willing to participate in a follow-up interview to explore in-depth their preference motives.

The interview was conducted on the same day as completion of the questionnaires of the previous study or at the next follow-up appointment (with a maximum delay of 6 months), or by phone, with a mean delay of 51.0 (SD 68.3) days. This delay was chosen to limit patient burden and seemed reasonable as no technology-based PA interventions were offered to the patients during this period. These 1-on-1 interviews were conducted in a specialized obesity center and organized in conjunction with outpatient visits or by phone by FH between June and December 2019. The descriptions of the interventions were presented again at the beginning of the interviews, which lasted a mean 17.3 (SD 5.2) minutes. Participants were asked to provide demographic data including (1) year of birth, (2) sex, (3) marital status, (4) educational level, and (5) self-reported height (m) and weight (kg) used for calculating the BMI (kg/m²).

Four researchers analyzed the data; 2 researchers were specialized in psychology and ergonomic sciences (FH and PT) and 2 were from the fields of exercise psychology and social psychology (MH and FA-L). Interviews were audio-recorded and transcribed verbatim. Interviewing ended when theoretical saturation was reached at the general level for all technology-based PA interventions combined [26]. Theoretical saturation was a guiding principle classically used to assess sample adequacy in qualitative research. Recently, a systematic review of empirical tests showed that 9-17 interviews reach saturation and transcribed verbatim. Interviewing ended when theoretical saturation was reached at the general level for all technology-based PA interventions combined [26]. Theoretical saturation was a guiding principle classically used to assess sample adequacy in qualitative research. Recently, a systematic review of empirical tests showed that 9-17 interviews reach saturation and transcribed verbatim. Interviewing ended when theoretical saturation was reached at the general level for all technology-based PA interventions combined [26].

The study was conducted in accordance with the Helsinki agreement of all researchers. As the UTAUT2 dimensions are research validity [34]. Category labels were refined with the rest of the data coding was then shared out between PT and FH read the units of meaning several times to become familiar with the data. They coded the units deductively into the main dimensions of the UTAUT2 model [19]: (1) performance expectancy, (2) effort expectancy, (3) social influence, (4) facilitating conditions, (5) hedonic motivation, (6) price value, (7) habit, and (8) behavioral intentions. They then determined the subcategories of the UTAUT2 dimensions inductively. Units of meaning that were not relevant for the UTAUT2 dimensions were organized into emergent new categories and subcategories. The 2 researchers independently coded 30% of the data (ie, 8 interviews out of 26) and obtained 94.02% agreement (ie, 1132 units of meaning were coded identically out of 1204 units). They then shared their coding and discussed any diverging results until agreement was reached. The rest of the data coding was then shared out between PT and FH. In the fourth step, FA-L and MH reviewed the categories and codes as “disinterested peers” to strengthen the qualitative research validity [34]. Category labels were refined with the agreement of all researchers. As the UTAUT2 dimensions are

**Interview Guide**

The interview guide was mainly based on the constructs of the UTAUT2 model [19]. A pilot interview enabled us to reformulate some of the questions and focus the interview on the preferred technology. The final guide comprised 4 parts: (1) presentation of the descriptions of the technology-based PA interventions and confirmation of the ranked preferences, (2) exploration of the reasons for the ranking, (3) application of the UTAUT2 dimensions to the preferred technology and comparison with the other interventions, and (4) exploration of other factors that could influence acceptability (Multimedia Appendix 2).

**Qualitative Analysis**

Qualitative analysis was conducted in several steps according to the qualitative research guidelines [26,30-32]. In the first step, PT and FH determined the segmentation procedure based on Strijbos et al [33] independently of the coding categories of our study. Data were segmented based on punctuation and subdivided when a segment included several units of meaning; conditional relations constituted 1 segment, and sentences left pending and speech tics were excluded. In the second step, PT and FH read the units of meaning several times to become familiar with the data. They coded the units deductively into the main dimensions of the UTAUT2 model [19]: (1) performance expectancy, (2) effort expectancy, (3) social influence, (4) facilitating conditions, (5) hedonic motivation, (6) price value, (7) habit, and (8) behavioral intentions. They then determined the subcategories of the UTAUT2 dimensions inductively. Units of meaning that were not relevant for the UTAUT2 dimensions were organized into emergent new categories and subcategories. The 2 researchers independently coded 30% of the data (ie, 8 interviews out of 26) and obtained 94.02% agreement (ie, 1132 units of meaning were coded identically out of 1204 units). They then shared their coding and discussed any diverging results until agreement was reached. The rest of the data coding was then shared out between PT and FH. In the fourth step, FA-L and MH reviewed the categories and codes as “disinterested peers” to strengthen the qualitative research validity [34]. Category labels were refined with the agreement of all researchers. As the UTAUT2 dimensions are preferred active video game as their second choice (other second choices: n=41 preferred mobile app and n=30 preferred telehealth); and (3) 63.9% (n=85) preferred telehealth as their third choice (other third choices: n=29 preferred active video game and n=19 preferred mobile app). As many as 26 of these women (preference choice: active video game, n=10; mobile app, n=9; and telehealth, n=7) volunteered to participate in this study. This subsample was not representative of the previous 133 participants in terms of preferences (ie, as a first choice, n=73, 54.9%, chose mobile app; n=42, 31.5%, chose active video game, and n=18, 13.5%, chose telehealth). As the objective of the study was to explore patient preference motives, we preferentially conducted the interviews with a view to balancing the number of participants preferring each of the technology-based PA interventions until theoretical saturation was achieved.

**Participants**

**Inclusion Criteria**

Inclusion criteria were the following: (1) women residing in France, (2) between 18 and 40 years of age, (3) having undergone bariatric surgery at least two months earlier, (4) with care received in the south of France, (5) without PA limitation, and (6) speaking French fluently. We focused on women because they undergo bariatric surgery more often than men and make up 82% of those undergoing this surgery in France [28]. Moreover, women undergo bariatric surgery at a younger age [29] and are more prone to physical inactivity and sedentary behaviors [11] than men. We restricted the inclusion criteria to young women to ensure sample homogeneity and to avoid confounding by UTAUT2 moderators such as age and sex.

We had performed an earlier quantitative study with patients with obesity about the acceptability of 3 technology-based PA interventions [17]. Among the 133 eligible participants, (1) 54.9% (n=73) preferred mobile app as their first choice (other first choices: n=42 preferred active video game and n=18 preferred telehealth); (2) 46.6% (n=62) preferred active video game as their second choice (other second choices: n=41 preferred mobile app and n=30 preferred telehealth); and (3) 63.9% (n=85) preferred telehealth as their third choice (other third choices: n=29 preferred active video game and n=19 preferred mobile app). As many as 26 of these women (preference choice: active video game, n=10; mobile app, n=9; and telehealth, n=7) volunteered to participate in this study. This subsample was not representative of the previous 133 participants in terms of preferences (ie, as a first choice, n=73, 54.9%, chose mobile app; n=42, 31.5%, chose active video game, and n=18, 13.5%, chose telehealth). As the objective of the study was to explore patient preference motives, we preferentially conducted the interviews with a view to balancing the number of participants preferring each of the technology-based PA interventions until theoretical saturation was achieved.

**Interview Guide**

The interview guide was mainly based on the constructs of the UTAUT2 model [19]. A pilot interview enabled us to reformulate some of the questions and focus the interview on the preferred technology. The final guide comprised 4 parts: (1) presentation of the descriptions of the technology-based PA interventions and confirmation of the ranked preferences, (2) exploration of the reasons for the ranking, (3) application of the UTAUT2 dimensions to the preferred technology and comparison with the other interventions, and (4) exploration of other factors that could influence acceptability (Multimedia Appendix 2).

**Qualitative Analysis**

Qualitative analysis was conducted in several steps according to the qualitative research guidelines [26,30-32]. In the first step, PT and FH determined the segmentation procedure based on Strijbos et al [33] independently of the coding categories of our study. Data were segmented based on punctuation and subdivided when a segment included several units of meaning; conditional relations constituted 1 segment, and sentences left pending and speech tics were excluded. In the second step, PT and FH read the units of meaning several times to become familiar with the data. They coded the units deductively into the main dimensions of the UTAUT2 model [19]: (1) performance expectancy, (2) effort expectancy, (3) social influence, (4) facilitating conditions, (5) hedonic motivation, (6) price value, (7) habit, and (8) behavioral intentions. They then determined the subcategories of the UTAUT2 dimensions inductively. Units of meaning that were not relevant for the UTAUT2 dimensions were organized into emergent new categories and subcategories. The 2 researchers independently coded 30% of the data (ie, 8 interviews out of 26) and obtained 94.02% agreement (ie, 1132 units of meaning were coded identically out of 1204 units). They then shared their coding and discussed any diverging results until agreement was reached. The rest of the data coding was then shared out between PT and FH. In the fourth step, FA-L and MH reviewed the categories and codes as “disinterested peers” to strengthen the qualitative research validity [34]. Category labels were refined with the agreement of all researchers. As the UTAUT2 dimensions are
defined as degrees, PT and MH specified independently for each participant whether the category cited was perceived positively, negatively, or neutrally. The authors obtained 93.74% agreement (ie, 494 codes were perceived identically out of 527 codes) and resolved disagreements by consensus. They then counted the number of participants who reported each category for each technology-based PA intervention. As a final step, relevant and short extract examples were identified and selected with the agreement of all researchers.

Table 1. Sociodemographic characteristics (n=26)

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>32.9 (5.5)</td>
</tr>
<tr>
<td>Body mass index (kg/m²), mean (SD)</td>
<td>30.1 (6.5)</td>
</tr>
<tr>
<td>Education (years), n (%)</td>
<td></td>
</tr>
<tr>
<td>&lt;12</td>
<td>10 (38.5)</td>
</tr>
<tr>
<td>12</td>
<td>9 (34.6)</td>
</tr>
<tr>
<td>14-15</td>
<td>7 (26.9)</td>
</tr>
<tr>
<td>≥17</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Professional status, n (%)</td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>20 (76.9)</td>
</tr>
<tr>
<td>Unemployed</td>
<td>5 (19.2)</td>
</tr>
<tr>
<td>Student</td>
<td>1 (3.8)</td>
</tr>
<tr>
<td>Marital status, n (%)</td>
<td></td>
</tr>
<tr>
<td>Single or never married</td>
<td>10 (38.5)</td>
</tr>
<tr>
<td>Married or in a civil union</td>
<td>12 (46.2)</td>
</tr>
<tr>
<td>Divorced or widowed</td>
<td>4 (15.4)</td>
</tr>
</tbody>
</table>

Qualitative Analysis

Overview

Units of meaning for each technology-based PA intervention were coded deductively into the main dimensions of the UTAUT2 model. Then, subcategories of the UTAUT2 dimensions were determined inductively. Codes that were not relevant for the UTAUT2 dimensions were organized into emergent new categories and subcategories (Table 2). The way each participant perceived the different acceptability categories and subcategories is reported in Multimedia Appendix 3. These perceptions are also summarized in Table 2.
<table>
<thead>
<tr>
<th>Categories and subcategories</th>
<th>Active video game</th>
<th>Mobile app</th>
<th>Telehealth</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n (%); valence&lt;sup&gt;a&lt;/sup&gt;</td>
<td>n (%); valence</td>
<td>n (%); valence</td>
</tr>
<tr>
<td><strong>UTAUT2 constructs</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance expectancy</td>
<td>19 (73.1)</td>
<td>19 (73.1)</td>
<td>17 (65.4)</td>
</tr>
<tr>
<td>Adequacy of PA&lt;sup&gt;c&lt;/sup&gt;</td>
<td>8 (30.8); 5 (-), 1 (+), 2 (±)</td>
<td>8 (30.8); 2 (-), 4 (+), 2 (±)</td>
<td>12 (46.2); 12 (+)</td>
</tr>
<tr>
<td>Engagement and sustainability of PA</td>
<td>16 (61.5); 1 (-), 14 (+), 1 (±)</td>
<td>12 (46.2); 2 (-), 9 (+), 1 (±)</td>
<td>14 (53.8); 1 (-), 13 (+)</td>
</tr>
<tr>
<td>PA management support</td>
<td>3 (11.5); 1 (-), 2 (+)</td>
<td>13 (50.0); 10 (+), 3 (±)</td>
<td>4 (15.4); 4 (+)</td>
</tr>
<tr>
<td>Effort expectancy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effort required by PA</td>
<td>11 (42.3)</td>
<td>14 (53.8)</td>
<td>5 (19.2)</td>
</tr>
<tr>
<td>Effort required by the technology</td>
<td>8 (30.8); 2 (-), 4 (+), 2 (±)</td>
<td>14 (53.8); 5 (-), 9 (+)</td>
<td>4 (15.4); 4 (+)</td>
</tr>
<tr>
<td>Social influence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Others’ perceptions on the technology-based PA interventions</td>
<td>8 (30.8); 1 (-), 5 (+), 2 (±)</td>
<td>10 (38.5); 8 (+), 2 (±)</td>
<td>4 (15.4); 2 (+), 2 (±)</td>
</tr>
<tr>
<td>Others’ uses of the technology-based PA interventions</td>
<td>6 (23.1); 1 (-), 4 (+), 1 (±)</td>
<td>3 (11.5); 3 (+)</td>
<td>1 (3.8); 1 (+)</td>
</tr>
<tr>
<td>Facilitating conditions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anytime and anywhere usage</td>
<td>22 (84.6)</td>
<td>24 (92.3)</td>
<td>20 (76.9)</td>
</tr>
<tr>
<td>Available material resources</td>
<td>17 (65.4); 7 (-), 6 (+), 4 (±)</td>
<td>22 (84.6); 2 (-), 18 (+), 2 (±)</td>
<td>18 (69.2); 11 (-), 5 (+), 2 (±)</td>
</tr>
<tr>
<td>Technological knowledge</td>
<td>11 (42.3); 2 (-), 5 (+), 4 (±)</td>
<td>11 (42.3); 10 (+), 1 (±)</td>
<td>9 (34.6); 3 (-), 5 (+), 1 (±)</td>
</tr>
<tr>
<td>Available human assistance</td>
<td>4 (15.4); 1 (-), 3 (+)</td>
<td>2 (7.7); 2 (+)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Hedonic motivation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Usage pleasure</td>
<td>19 (73.1); 2 (-), 16 (+), 1 (±)</td>
<td>9 (34.6); 1 (-), 8 (+)</td>
<td>6 (23.1); 3 (-), 3 (+)</td>
</tr>
<tr>
<td>Usage interest</td>
<td>14 (53.8); 8 (-), 6 (+)</td>
<td>5 (19.2); 5 (-)</td>
<td>3 (11.5); 2 (-), 1 (±)</td>
</tr>
<tr>
<td>Price value</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Willingness to pay</td>
<td>13 (50.0); 4 (-), 5 (+), 4 (±)</td>
<td>10 (38.5); 1 (-), 7 (+), 2 (±)</td>
<td>8 (30.8); 4 (-), 3 (+), 1 (±)</td>
</tr>
<tr>
<td>Financial savings</td>
<td>0 (0)</td>
<td>1 (3.8); 1 (+)</td>
<td>2 (7.7); 2 (+)</td>
</tr>
<tr>
<td>Habit</td>
<td>19 (73.1)</td>
<td>18 (69.2)</td>
<td>13 (50.0)</td>
</tr>
<tr>
<td>Use of PA technology</td>
<td>15 (57.7); 6 (-), 9 (+)</td>
<td>15 (57.7); 6 (-), 9 (+)</td>
<td>12 (46.2); 6 (-), 4 (+), 2 (±)</td>
</tr>
<tr>
<td>Use of similar technology</td>
<td>10 (38.5); 4 (-), 5 (+), 1 (±)</td>
<td>8 (30.8); 2 (-), 5 (+), 1 (±)</td>
<td>3 (11.5); 2 (-), 1 (±)</td>
</tr>
<tr>
<td>Emerging categories</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other motivational factors</td>
<td>18 (69.2)</td>
<td>10 (38.5)</td>
<td>20 (76.9)</td>
</tr>
<tr>
<td>Motivation to be related to others</td>
<td>14 (53.8); 4 (-), 10 (+)</td>
<td>9 (34.6); 7 (-), 2 (+)</td>
<td>19 (73.1); 4 (-) 13 (+), 2 (±)</td>
</tr>
<tr>
<td>Motivation for competition</td>
<td>6 (23.1); 6 (+)</td>
<td>0 (0)</td>
<td>1 (3.8); 1 (+)</td>
</tr>
<tr>
<td>Motivation for health</td>
<td>1 (3.8); 1 (+)</td>
<td>3 (11.5); 3 (+)</td>
<td>4 (15.4); 4 (+)</td>
</tr>
<tr>
<td>Other characteristics</td>
<td>2 (7.7)</td>
<td>4 (15.4)</td>
<td>4 (15.4)</td>
</tr>
<tr>
<td>Perceived reliability</td>
<td>0 (0)</td>
<td>3 (11.5); 3 (-)</td>
<td>1 (3.8); 1 (-)</td>
</tr>
<tr>
<td>Intimacy preservation</td>
<td>2 (7.7); 2 (+)</td>
<td>1 (3.8); 1 (+)</td>
<td>4 (15.4); 3 (-), 1 (±)</td>
</tr>
<tr>
<td>Distraction by other technology features</td>
<td>0 (0)</td>
<td>3 (11.5); 3 (+)</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Valence is the number of participants who expressed the different acceptability categories and subcategories negatively (-), positively (+), or neutrally (±).

<sup>b</sup>Categories in italics were summed based on a count of individual participants who mentioned at least one of the subcategories.

<sup>c</sup>PA: physical activity.
The Choice of Preferred Technology-Based PA Interventions and Behavioral Intentions to Use Them

Among the 26 participants, 10 indicated during the interview their preference for active video game, 9 for mobile app, and 7 for telehealth. Between the time they agreed to participate in this study and the interview, 1 participant changed her choice and preferred mobile app instead of telehealth (P17). All participants who preferred a technology-based PA intervention (ie, active video game, n=10; mobile app, n=10; telehealth, n=6) expressed a behavioral intention to use it. In addition, among the participants who preferred another technology-based PA intervention (ie, second and third choices), (1) 3 expressed low behavioral intention for active video game (ie, P8, P10, and P25), (2) 1 expressed high behavioral intention for active video game (ie, P22), (3) 1 expressed low behavioral intention for mobile app (ie, P20), (4) 3 expressed low behavioral intention for telehealth (ie, P16, P17, and P22), and (5) 4 expressed high behavioral intention for telehealth (ie, P7, P12, P14, and P26). The following excerpts illustrate these results:

(mobile app) I'll use it...well after...yeah, I think, all the time [P5]
(telehealth) ah, but if I have it at home, I'll do it all the time [P19]
I can't, I can't say to myself, well I'm going to turn on a video game to do some sports [P8]
but what is certain is that I'm not interested in telehealth [P22]

UTAUT2 Constructs

Performance Expectancy

This category of the UTAUT2 referred to the degree to which the participants believed that the technology would be useful to them in doing PA. Three subcategories emerged for the 3 technology-based PA interventions: (1) adequacy of PA, (2) engagement and sustainability of PA, and (3) PA management support. Among the participants who mentioned the adequacy of PA for active video game (n=8), most perceived it to be of low adequacy; for example, “there is no real contact, or the descriptions are badly done, or something like that” [P2].

For mobile app (n=8), perceptions were quite good about the adequacy of PA. For telehealth, adequacy was perceived as high among the 12 participants who mentioned this subcategory; for example,

to see if we're doing the right things, if we're doing the exercise correctly, so that we're not doing anything and everything [P2]

Perceptions of the technologies to engage and sustain PA throughout a session or over the long term were generally positive for active video game (n=16), mobile app (n=12), and telehealth (n=14), as highlighted by the following quotes:

but maybe to start, you know, as a first step to get back into sports, it's maybe more interesting to start with the video game [P26]

(telehealth) even if it’s on the computer, it motivates us, it pushes us a little bit to improve, to go a little further [P24]

For active video game (n=3), mobile app (n=13), and telehealth (n=4), participants perceived these technologies as mostly helping them to manage and monitor their PA, as noted by one of the participants: “that we can see our progress on the application.” [P6]

Effort Expectancy

This category of the UTAUT2 referred to the degree to which the participants believed that the technology would be easy to use for PA. First, participants mentioned the perceived ease of use in relation to the physical effort involved in PA. For active video game, participants (n=6) perceived this to a mixed degree as illustrated by the following quotes:

if, for example, he asks me to jump, I'll jump, but uh, my knee will hurt [P4]
precisely when it's a video game, there are several levels. [P16]

For mobile app (n=4) and telehealth (n=2), the effort involved in PA was perceived to be low and adapted to their capacities; for example, mobile app was perceived as “adapted to each level, so it's good for making progress” [P7]. Second, participants perceived the effort required by the technology as low (ie, active video game, n=8; mobile app, n=14; telehealth, n=4), which refers to the concept of the usability of the technology-based PA interventions. One participant stated as follows: “(telehealth) one click and it starts up by itself, it seems simple to me” [P8].

Social Influence

This category of the UTAUT2 referred to the degree to which the participants perceived that significant others believed they should use the technology-based PA intervention to do PA. Two subcategories emerged for the interventions: (1) others’ perceptions of the technology (ie, subjective norms), and (2) others’ uses of the technology (ie, descriptive norms). For all the technology-based PA interventions, others’ perceptions of the technology (ie, active video games, n=8; mobile apps, n=10; telehealth, n=4) and others’ uses of it (ie, active video games, n=6; mobile apps, n=3; telehealth, n=1) were mostly perceived positively. For example, participants stated:

everyone plays these games at least a little bit so they would find it normal [P13] (telehealth) perhaps there would be some curious ‘ah, but how does it work? Can I try to do a session with you?’ [P20]
(mobile app) maybe they would even use it, who knows [P6]

Facilitating Conditions

This category of the UTAUT2 referred to the participants’ perceptions of the resources and support available to them while using the technology-based PA interventions. Four subcategories emerged: (1) anytime and anywhere usage, (2) available material resources, (3) technological knowledge, and (4) available human assistance. Participants perceived mobile app (n=22) to be usable anytime and anywhere, as noted by P23: “an application you...
can do it whenever you want, so when you have some time,” whereas the perception of active video game (n=17) was more nuanced: “having to be at home to do it, it’s more restrictive” [P11]. For telehealth (n=18), participants mostly perceived it as usable to a limited extent: “having to keep a schedule could be complicated for me” [P1], except for those who preferred this technology-based intervention and perceived it as adapted to their lifestyle and allowing them to save transport time: “we’re going to be able to organize ourselves more easily according to, well, our daily lives, we’re not going to lose time in transportation” [P8].

For mobile app (n=11) and telehealth (n=9), the participants felt they had material resources available, as illustrated by the following quote:

(telehealth) I’ve got the smartphone on which I’ve got a webcam, I’ve got the computer with it so, um well, hardware-wise I’ll have everything [P8]

For active video game (n=11), the necessary equipment was not always available; for example, “video games you have to have the equipment, so sometimes you can’t have it” [P23].

Technological knowledge needed to use the technology-based PA interventions was cited to a lesser extent (ie, active video game, n=3; mobile app, n=4; telehealth, n=4), but mostly perceived positively, as this excerpt shows:

(telehealth) none because, although I’m not much of a TV person or anything, I know how to use computers, plug in, connect or whatever [P20]

Available human assistance for using technology-based PA interventions was reported positively for mobile app (n=2) and active video game (n=4), as illustrated by P1, “(active video game) by giving me time to do it, maybe do it with me,” but was not reported for telehealth.

Hedonic Motivation

This category of the UTAUT2 referred to the participants’ cognitive trade-offs between the perceived benefits of the technology-based PA interventions for doing PA and the estimated monetary cost of using them. Participants expressed a degree of willingness to pay to use the technology-based PA interventions, provided the price was not too high. Mobile app was considered to have an acceptable price by those who mentioned this subcategory (n=10), while opinions were more mixed for active video game (n=13) and telehealth (n=8). The following quotes illustrate this:

pay for the application, I wouldn’t mind to a certain extent [P2]

mobile app) if it’s in a gym or if it’s my phone, um...in the gym I say to myself, if I like it I’ll go, I’ll pay, so it would be the same [P17]

(active video game) we are not going to say that it’s within our reach [P14]

(telehealth) I know that, even if it would have to be paid for; and I know that I would be willing to pay the price [P10]

To a lesser extent, 3 participants also mentioned the financial savings with the technology-based PA interventions, especially mobile app (n=1) or telehealth (n=2), compared with the gym, as illustrated by P10: ‘I’m sure by telehealth and all of that, it would be much cheaper.”

Habit

This category of the UTAUT2 was extended from the original definition and referred to the previous use of the technology-based PA interventions or a similar technology. Thus, 2 subcategories emerged: (1) the use for PA of the technology described in the presentations or similar technology, and (2) the use of a similar technology for activities other than PA. For active video game and mobile app, participants described having rather a high use of similar technologies for PA (ie, active video game, n=15; mobile app, n=15) and for other activities (ie, active video game, n=10; mobile app, n=8), as illustrated by these quotes:
By contrast, for telehealth, the use of similar technologies for PA (n=12) or other activities (n=3) was rather perceived as low, as cited by P14 “in telehealth, since I’ve never tested it, so I don’t know.”

**Emerging Categories**

**Other Motivational Factors**

This category corresponded to the motivational determinants of technology-based PA intervention use that went beyond the motivational factors included in the hedonic motivation and performance expectancy constructs of the UTAUT2. Three other motivational factors emerged: (1) motivation to be related to others, (2) motivation for competition, and (3) motivation for health. The motivation to be related to others (ie, need for relatedness) referred to the motivation to use the technology-based PA interventions to be included in a group or to be connected with other people to do PA or with a coach. For active video game (n=14) and telehealth (n=19), participants perceived these technology-based PA interventions as a response to their need for relatedness:

(active video game) then I think that yeah, with an evening with friends or with children, it can be really nice. [P20]

(telehealth) the good thing is, if I remember, there was the possibility to be with a coach or with a group. [P23]

By contrast, mobile app (n=9) was mostly perceived as foreign to this need; for example,

the application I put it last because am being alone to do my sport is not very motivating. [P13]

Motivation for competition (ie, performance achievement goals) referred to the use of the technology-based PA interventions to measure oneself against others and compare oneself in a kind of competition. This subcategory was mainly mentioned for active video game (n=6); for example, P24 considered active video games as allowing “a little competition with people.” One participant (ie, P26) also mentioned this for telehealth. Motivation for health referred to the use of the interventions to improve physical capacities, lose weight, or avoid obesity relapse. This subcategory was cited positively for active video game (n=1), mobile app (n=3), and telehealth (n=4), as illustrated in the following quote: “(active video game) I think it can give me more...endurance, cardio” [P26].

**Other Characteristics**

This category corresponded to constructs that were not included in the UTAUT2 and were not related to motivational factors. Perceived reliability was sometimes perceived as low for mobile app (n=3) and telehealth (n=1), as noted by P7: “(mobile app) if it’s stuff that grinds uh, or bugs, well that’s annoying.”

Active video game (n=2) and mobile app (n=1) were perceived by some participants as preserving their intimacy because these technology-based PA interventions did not require them to expose themselves. By contrast, 3 participants perceived telehealth as exposing them; for example,

for the telehealth, uh...negative point is that sometimes, we don’t really want to show ourselves [P26]

One participant, however, considered that she was less exposed than in a gym:

and uh to do it at home without anyone around who can judge me like in a room or uh...look at me [P26]

Three participants mentioned that they might be distracted by other features on their smartphone instead of using the application for PA, as illustrated in the following excerpt:

when I pick up the smartphone, well immediately my games take over, I do something else, I go to answer the phone and then I make a phone call and finally I don’t do what I went to do on my phone [P11]

Beyond the study objectives, 2 participants perceived the proposed technology-based PA interventions as complementary (ie, the 3 interventions for P26 and mobile app and telehealth for P5).

**Discussion**

**Principal Findings**

The aim of this study was to examine the suitability of the UTAUT2 model for technology-based PA interventions in the context of bariatric surgery. To this end, we explored the reasons for preference for 1 of 3 interventions (ie, telehealth, active video game, and mobile app) to gain an in-depth insight into the factors contributing to behavioral intention to use the technology. Of the 26 participants, 10 chose active video game as their preferred technology-based PA intervention and 11 expressed a high behavior intention to use it, 10 preferred the mobile app and 10 intended to use it, and 6 chose telehealth and 10 intended to use it.

For active video game, the main positive factors mentioned by the participants were usage pleasure, engagement and sustainability of PA, and motivation to be related to others. By contrast, usage anytime/anywhere and motivation to be related to others were perceived more negatively. These specificities can serve as benchmarks for the development of future active video games targeting women in postbariatric surgery. For example, we recommend that the developers of these games stimulate usage pleasure, which could be achieved with less demanding physical exercises. For mobile app, the possibility to use it anytime and anywhere, the availability of material resources, and support for PA management were the most positively mentioned factors, while usage interest and motivation to be related to others were perceived less positively. According to these specificities, we could recommend short PA sessions or those based on everyday movements with little or no equipment. For telehealth, the adequacy of PA, engagement and sustainability of PA, and the motivation to be related to others were widely perceived positively, while telehealth was perceived as constraining for anytime and anywhere usage. We recommend that qualified professionals teach PA through this type of technology, with

---

https://humanfactors.jmir.org/2023/1/e42178

JMRI Hum Factors 2023 | vol. 10 | e42178 | p.460

(page number not for citation purposes)
some flexibility in booking slots and choice of extended hours. To the best of our knowledge, this study is the first to identify the most salient factors explaining the preferences of vulnerable people regarding technology-based PA interventions.

All the UTAUT2 constructs were broken down into subcategories specifically adapted to technology-based PA interventions in bariatric surgery, differing from other technologies used in chronic diseases. For example, facilitating conditions in diabetes mobile health (mHealth) self-management are broken down into technical support, support from the mHealth app itself, and health care professionals [24]. In our study, facilitating conditions in the technology-based PA interventions were broken down into anytime and anywhere usage, available material resources, technological knowledge, and available human assistance. Although some studies have used the UTAUT2 for technology-based PA interventions [35,36], to our knowledge this is the first study to characterize in-depth the concepts of the UTAUT2 model in this context for a vulnerable population. These findings validated the suitability of the UTAUT2 model in this context. However, future studies would be necessary to extend these results to other clinical contexts.

Our results showed that factors other than the constructs of the UTAUT2 model also emerged to characterize the acceptability of technology-based PA interventions. The UTAUT2 model combines several theories, such as the hierarchical model of intrinsic and extrinsic motivation [37]. The concept of performance expectancy integrates extrinsic motivation, and the concept of hedonic motivation integrates intrinsic motivation [38]. The UTAUT2 is a recognized theoretical framework for technology acceptability [20,21], which has been extended to several contexts [25]. However, few studies have considered the specificities of the acceptability of technology-based PA interventions in light of more contemporary sociocognitive models of motivation. Our results have been discussed in relation to the Self-Determination Theory (SDT) [39] and achievement goal theory [40]. In particular, motivation to be related to others corresponds to the need for relatedness, which is one of the basic psychological needs of the SDT. The extrinsic motivation, as cited by the participants (ie, motivation for health), referred to identified regulation among the 4 types of regulation of extrinsic motivation of the SDT. These results are in line with the findings of recent studies that have examined the relations between these theories and acceptability theories [41-45].

The examination of the relationships between the concepts of motivation to PA and the constructs of the UTAUT2 model in the context of technology-based PA interventions seems to be an emerging area of research that should be encouraged. As motivation toward PA has a higher degree of generality than motivation toward technology-based PA interventions (ie, performance expectancy and hedonic motivation), the SDT constructs could be positioned as antecedents of the UTAUT2 variables.

**Limitations**

Despite the several strengths of this study, some limitations must be acknowledged. The first limitation is related to the study design. The descriptions of the technology-based PA interventions provided general information and were relatively similar to avoid any bias to the presentation itself. As the descriptions were hypothetical, we cannot apply these results directly to similar real technology-based PA interventions available on the market. Although we conducted our qualitative analyses according to research guidelines [26,30-32] and reached theoretical saturation, the generalizability of our results may be questioned. First, our population was composed only of young women who underwent bariatric surgery. We can assume that young adults are rather familiar with technology. Second, those who agreed to participate in the interviews may have been more interested in technology-based PA interventions than the rest of the population. Third, there was no process for having the participants validate the results, such as member checking. Some of their responses may thus have been slightly overinterpreted.

Another type of limitation was related to our theoretical approach. The interviews were conducted within the framework of the UTAUT2 model, which means that the model constructs did not emerge naturally (ie, their frequency of citation is probably overestimated), unlike the other constructs, such as the motivational constructs. The relative weight of each of the factors in explaining behavioral intention to use technology-based PA interventions will have to be established in future studies, as will the relation with usage behavior, which was not measured in this study.

Face-to-face contact was minimized to lower the risk of virus transmission during the COVID-19 pandemic, which meant that telehealth was used extensively. As the interviews were conducted before the pandemic, perceptions about telehealth may have changed (eg, [46]).

**Conclusions**

The results showed that the UTAUT2 model is suitable for examining the acceptability of technology-based PA interventions in the context of bariatric surgery. All UTAUT2 constructs were broken down into subcategories specifically tailored to this context. The results also highlighted the most salient factors explaining the preferences of vulnerable individuals regarding several types of technology-based PA interventions. These results have important implications as they could be used as benchmarks for future technology development. Although the UTAUT2 model is an integrative model, other factors of acceptability were identified. Future studies must be conducted to better examine the causal relationship between the SDT and UTAUT2 constructs.

**Acknowledgments**

The authors express their sincere gratitude to all the volunteers who participated in this study. We also express our gratitude to the following for their aid in contacting the volunteers: The Specialized Center for Obesity of East Provence-Alpes Côte d’Azur
and the Nice University Hospital Center, the Specialized Center for Obesity of West Provence-Alpes Côte d’Azur and the Marseille University Hospital Center, the Nutrition Center of Pegomas, the Cérès Nutrition Center of Nice, and the Antibes Hospital Center. MH was supported by a PhD grant from the Région Sud Provence-Alpes Côte d’Azur, France, and co-supported by the association “Azur Sport Santé.” This work was supported by the French government and managed by the “Agence Nationale de la Recherche” as part of the UCA JEDI Future Investments project, reference number ANR-15-IDEX-01. The article processing fee was funded by the “Fédération de Recherche Interventions en Santé” (FRIS) and the “Maison des Sciences Humaines et Sociales Sud-Est” (MSHS Sud-Est) of the Université Côte d’Azur.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Original French and English translation of the written descriptions of three technology-based physical activity interventions with illustrations.

[DOCX File, 303 KB - humanfactors_v10i1e42178_app1.docx]

Multimedia Appendix 2
English translation of the interview guide.

[DOCX File, 25 KB - humanfactors_v10i1e42178_app2.docx]

Multimedia Appendix 3
Participants’ perceptions of the different acceptability categories and subcategories.

[DOCX File, 26 KB - humanfactors_v10i1e42178_app3.docx]

References


34. Johnson RB. Examining the validity structure of qualitative research. Education 1997;118(2):282-294 [FREE Full text]


Abbreviations
mHealth: mobile health
PA: physical activity
SDT: Self-Determination Theory
UTAUT2: Unified Theory of Acceptance and Use of Technology 2

©Pierre Thérouanne, Meggy Hayotte, Florent Halgand, Fabienne d'Arripe-Longueville. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 23.01.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Understanding the Subjective Experience of Long-term Remote Measurement Technology Use for Symptom Tracking in People With Depression: Multisite Longitudinal Qualitative Analysis

Katie M White, BSc; Erin Dawe-Lane, MSc; Sara Siddi, PhD; Femke Lamers, PhD; Sara Simblett, PhD; Gemma Riquelme Alacid, MSc; Alina Ivan, MSc; Inez Myin-Germeys, PhD; Josep Maria Haro, PhD; Carolin Oetzmann, MSc; Priya Popat, BSc; Aki Rintala, PhD; Elena Rubio-Abadal, PhD; Til Wykes, PhD; Claire Henderson, PhD; Matthew Hotopf, PhD; Faith Matcham, PhD

Department of Psychological Medicine, King’s College London, London, United Kingdom
Department of Psychology, King’s College London, London, United Kingdom
Parc Sanitari Sant Joan de Dèu, Fundació Sant Joan de Dèu, CIBERSAM, Universitat de Barcelona, Barcelona, Spain
Department of Psychiatry, Amsterdam University Medical Center, Vrije Universiteit, Amsterdam, Netherlands
Center for Contextual Psychiatry, Department of Neurosciences, UK Leuven, Leuven, Belgium
Health Service & Population Research Department, King’s College London, London, United Kingdom
School of Psychology, University of Sussex, Falmer, Sussex, United Kingdom

Corresponding Author:
Katie M White, BSc
Department of Psychological Medicine
King’s College London
Institute of Psychiatry, Psychology and Neuroscience
16 de Crespigny Park
London, SE5 8AB
United Kingdom
Phone: 44 7850684847
Email: katie.white@kcl.ac.uk

Abstract

Background: Remote measurement technologies (RMTs) have the potential to revolutionize major depressive disorder (MDD) disease management by offering the ability to assess, monitor, and predict symptom changes. However, the promise of RMT data depends heavily on sustained user engagement over extended periods. In this paper, we report a longitudinal qualitative study of the subjective experience of people with MDD engaging with RMTs to provide insight into system usability and user experience and to provide the basis for future promotion of RMT use in research and clinical practice.

Objective: We aimed to understand the subjective experience of long-term engagement with RMTs using qualitative data collected in a longitudinal study of RMTs for monitoring MDD. The objectives were to explore the key themes associated with long-term RMT use and to identify recommendations for future system engagement.

Methods: In this multisite, longitudinal qualitative research study, 124 semistructured interviews were conducted with 99 participants across the United Kingdom, Spain, and the Netherlands at 3-month, 12-month, and 24-month time points during a study exploring RMT use (the Remote Assessment of Disease and Relapse-Major Depressive Disorder study). Data were analyzed using thematic analysis, and interviews were audio recorded, transcribed, and coded in the native language, with the resulting quotes translated into English.

Results: There were 5 main themes regarding the subjective experience of long-term RMT use: research-related factors, the utility of RMTs for self-management, technology-related factors, clinical factors, and system amendments and additions.

Conclusions: The subjective experience of long-term RMT use can be considered from 2 main perspectives: experiential factors (how participants construct their experience of engaging with RMTs) and system-related factors (direct engagement with the technologies). A set of recommendations based on these strands are proposed for both future research and the real-world implementation of RMTs into clinical practice. Future exploration of experiential engagement with RMTs will be key to the successful use of RMTs in clinical care.
remote measurement; technology; qualitative; engagement; telehealth; depression; mental health; mobile phone

**Introduction**

**Background**

Depressive disorders, characterized by periods of persistent low mood and anhedonia, are the third leading cause of disability worldwide [1]. Major depressive disorder (MDD) is characterized by a longitudinal trajectory of relapse and remission [2]. The economic burden of MDD is currently estimated at US $326 billion [3], with high recurrence associated with increased comorbidity burden and health care resource use [4]. Traditional assessment of MDDs is limited in its ability to detect moment-by-moment symptom changes because it relies on retrospective questionnaires completed at sporadic time points, is prone to recall bias, and is often only undertaken at the point of relapse [5]. Working toward the timely diagnosis and treatment of MDD remains an urgent priority [5].

Novel remote measurement technologies (RMTs) have the potential to become an asset for chronic disease management. Multiparametric RMT systems can provide real-time, longitudinal symptom tracking by combining active symptom reporting via smartphone apps (active RMT) with physiological and behavioral wearable sensor data (passive RMT) [6]. Continuous data can be collected on mood variability [7], sociability [8], physical activity [9], cognition [10], speech acoustics [11], and sleep [12]. Integration of RMT data into MDD care may help to more accurately assess, monitor, and predict depressive symptom trajectories, ultimately enabling personalized interventions [13].

The promise of remote tracking in MDD depends almost entirely on user engagement. Engagement with mobile health (mHealth) technologies comprises the initial and sustained active use of a device [14]. High engagement with RMTs is imperative given the high-frequency data needed to identify symptom patterns and changes over time. Several systematic reviews have highlighted the heterogeneity of engagement metrics reported in remote tracking studies [15-17]. The Remote Assessment of Disease and Relapse-Major Depressive Disorder (RADAR-MDD) study is currently the largest multisite longitudinal study of a multiparametric RMT system for tracking depression [6]. The RADAR-MDD study has recently reported promising engagement, both in terms of initial recruitment rates [18] and sustained retention and data availability [19] over a 2-year follow-up of 623 participants across 3 European sites (United Kingdom, Spain, and the Netherlands). A large proportion of participants (79.8%) completed follow-up, and approximately 50% of the participants had >76% data completion for passive data streams [19].

When evaluating engagement, an understanding of the subjective experience of using RMTs should complement objective data completion statistics [17]. Subjective engagement with mHealth technologies can be understood as an experiential construct of what it feels like [20]. Exploring subjective engagement with RMTs provides a richer insight into system usability and perceived utility of, and satisfaction with, the technology [17]. The drivers for sustained user engagement with RMT systems, which, in contrast to typical mHealth technologies, require long periods of use for little direct rewards or intervention [21], are currently unknown.

Several studies have qualitatively explored subjective engagement with RMTs for depression. A multisite exploration of the perceived barriers and facilitators to RMT use by Simblett et al [22] informed the design of the RADAR-MDD study. Functional (technological convenience, accessibility, and intrusiveness) and nonfunctional (user cognition, perceived rewards) factors influenced patients when considering remote symptom tracking [22]. These findings have been replicated across patient and physician perspectives [23-25]. Two systematic reviews [26,27] on broader mHealth technologies for depression explored the experiences of participants’ actual use for up to 1 year. Factors such as lower symptom severity, perceived usefulness of the technology, lower privacy concerns, lack of technical issues, and access to responsive personal support were associated with enhanced motivation to engage with technologies [26,27]. A handful of studies have also suggested the beneficial effects of symptom monitoring, including increased self-awareness [28], adaptation of self-management strategies [29], and access to a “safety net” of support [30]. However, these studies typically use hypothetical scenarios or evaluate short-term system use. As a result, little is known about the subjective experience of long-term, real-world use of RMTs.

**Objective**

This study aims to understand the subjective experience of long-term engagement with RMTs for monitoring depression symptoms. It uses qualitative data from the RADAR-MDD study as an example of sustained RMT use across a 2-year follow-up period. This study builds on previous qualitative work by Simblett et al [22] on perceived barriers to and facilitators of intended RMT use in depression, providing a comparison with user experiences over 2 years of sustained engagement. Our objectives were (1) to explore key themes associated with long-term RMT use and (2) to identify recommendations for future system engagement. The findings will complement the objective engagement data and provide a basis for further promotion of engagement with RMTs for symptom tracking in research and clinical practice.

**Methods**

**Design**

This study used a multisite longitudinal qualitative research [31] approach with thematic analysis. Semistructured interviews were conducted with participants at 3-, 12-, and 24-month time points at 3 RADAR-MDD sites: King’s College London (London, United Kingdom), Centro de Investigación Biomédica (London, United Kingdom), and St. Mary’s Hospital, London (London, United Kingdom). The aim was to understand the subjective experience of remote tracking in MDD and to identify key themes associated with sustained RMT use.
en Red (Barcelona, Spain), and Amsterdam University Medical Centre (Amsterdam, the Netherlands). The design of the interview topic guide was informed by recent work on the barriers to and facilitators of RMT use in those living with depression [16,22].

Procedure
The RADAR-MDD study used the RADAR-base system [32] for data collection. The study active RMT smartphone app delivered fortnightly validated mood and self-esteem questionnaires and 6-weekly, high-frequency experience sampling methodology (ESM) questionnaires on current state, cognitive games, and a speech task. The study passive RMT smartphone app collected passive data on ambient noise and light, Bluetooth connection, and GPS location. Participants were provided with a wearable device, the Fitbit Charge (Fitbit Inc), measuring their step count, sleep, and physical activity. Further information on the RADAR-MDD procedure is available in the protocol paper by Matcham et al [6].

Eligibility criteria for inclusion in this study were (1) current participation in RADAR-MDD (full eligibility criteria provided in the study by Matcham et al [6]) and (2) willingness to participate in a 1:1 interview with a researcher discussing their experiences of the study. Participants provided written informed consent for the interviews as part of their RADAR-MDD study participation.

The interviews were managed by the research team lead at each site. Participants were recruited using convenience sampling at each time point to maximize data collection. Interviews were face-to-face (at the respective research site) or via telephone or video call (United Kingdom and the Netherlands only). All interviewers were female and part of the participant-facing research team. Face-to-face interviews were not conducted during the COVID-19 pandemic lockdown. Participants were reimbursed for relevant travel costs and paid per interview (£10 or €10 [US $1.2]).

The interviews were semistructured using open-ended questions, designed to elicit discussions around using the study technology in daily life (Multimedia Appendix 1). The content of each topic guide reflected the expected differences between time points. For example, the 3-month interview guide focused on immediate problem-solving and troubleshooting, while later interviews included data sharing.

The topic guides were translated from English into Spanish and Dutch, and interviews were conducted by native speakers at each site. The interviews lasted between 30 and 60 minutes and were conducted between February 2018 and April 2021.

Ethics Approval
The semistructured interviews were approved by the ethics committee of RADAR-MDD [6]. Ethical approvals for conducting the study were obtained from Camberwell St Giles Research Ethics Committee (reference: 17/LO/1154) in London, from Clinical Research Ethics Committee Fundacio Sant Joan de Déu (CI: PIC-128-17) in Barcelona, and from Medische Ethische Toetsingscommissie VUms (2018.012–NL63557.029.17) in the Netherlands.

Data Analysis Strategy
The interviews were audio recorded and transcribed verbatim. A preliminary coding framework was developed in English based on previous findings of barriers to and facilitators of RMT use in hypothetical scenarios [22]. All sites first coded example interviews for a cross-site consistency check and a discussion on revisions to the coding framework, accounting for novel codes. Each site then proceeded to recode all interviews in the native language using NVivo software (version 12; QSR International [33]) according to the final coding framework (Multimedia Appendix 2 provides a comparison of the preliminary and final coding framework). The coding was performed by independent researchers at each site. Each site sent coded NVivo data sets to the London site, with all quotes translated into English by a third-party translator briefed on the study topic [34]. The data were stored on a secure server at the London site.

Multisite data were merged into one data set and thematic maps for 3-month, 12-month, and 24-month time points were developed by 3 researchers (KW, EDL, and PP), identifying key themes and subthemes. To align with previous longitudinal qualitative research work [31], data are presented not as a longitudinal narrative but as contributing to each theme.

Results
Participant Characteristics
A total of 124 interviews with 99 participants were conducted across 3 sites. Of these 124 interviews, 40 (32.2%) interviews were conducted at the 3-month time point (15/40, 38% in United Kingdom; 15/40, 38% in Spain; and 10/40, 25% in the Netherlands), 42 (33.9%) at the 12-month time point (16/42, 38% at United Kingdom; 16/42, 38% at Spain; 10/42, 24% at the Netherlands), and 42 (33.9%) at the 24-month time point (15/42, 36% at United Kingdom; 16/42, 38% at Spain; 11/42, 26% at the Netherlands). A total of 17 participants took part in an interview at 2 time points; 4 participants were interviewed across all 3 time points. Participant characteristics according to time points are shown in Table 1.
Table 1. Participant characteristics by interview time point.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>3-month (n=40)</th>
<th>12-month (n=42)</th>
<th>24-month (n=42)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site, n</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>United Kingdom</td>
<td>15</td>
<td>16</td>
<td>15</td>
</tr>
<tr>
<td>Spain</td>
<td>15</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>the Netherlands</td>
<td>10</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>44.6 (12.1)</td>
<td>49.4 (13.5)</td>
<td>51.9 (15.0)</td>
</tr>
<tr>
<td>Female, n (%)</td>
<td>30 (75)</td>
<td>32 (76)</td>
<td>29 (69)</td>
</tr>
<tr>
<td><strong>Depression severity category</strong>, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>4 (10)</td>
<td>3 (7)</td>
<td>5 (12)</td>
</tr>
<tr>
<td>Mild</td>
<td>7 (18)</td>
<td>5 (12)</td>
<td>5 (12)</td>
</tr>
<tr>
<td>Moderate</td>
<td>10 (25)</td>
<td>13 (31)</td>
<td>7 (17)</td>
</tr>
<tr>
<td>Severe</td>
<td>7 (18)</td>
<td>10 (24)</td>
<td>6 (14)</td>
</tr>
<tr>
<td>Very severe</td>
<td>11 (28)</td>
<td>9 (21)</td>
<td>5 (12)</td>
</tr>
<tr>
<td>Not reported</td>
<td>1 (3)</td>
<td>2 (5)</td>
<td>14 (33)</td>
</tr>
<tr>
<td><strong>Anxiety severity category</strong>, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>7 (18)</td>
<td>5 (12)</td>
<td>7 (17)</td>
</tr>
<tr>
<td>Mild</td>
<td>7 (18)</td>
<td>10 (24)</td>
<td>8 (19)</td>
</tr>
<tr>
<td>Moderate</td>
<td>12 (30)</td>
<td>13 (31)</td>
<td>7 (17)</td>
</tr>
<tr>
<td>Severe</td>
<td>13 (33)</td>
<td>12 (329)</td>
<td>5 (12)</td>
</tr>
<tr>
<td>Not reported</td>
<td>1 (3)</td>
<td>2 (4)</td>
<td>15 (36)</td>
</tr>
</tbody>
</table>

*a Measured as the Inventory of Depressive Symptomatology-Self Report total score nearest to the interview time for each participant. None=0-13, mild=14-25, moderate=26-38, severe=39-48, and very severe=49-84.

*b Measured as the Generalized Anxiety Disorder-7 item total score nearest to the interview time for each participant. None=0-5, mild=6-10, moderate=11-15, and severe=16-21.

Themes

This study aimed to explore the subjective experience of long-term engagement with RMTs over a 2-year follow-up period. We present our results under five themes: (1) research-related factors, (2) the utility of RMTs for self-management, (3) technology-related factors, (4) clinical factors, and (5) system amendments and additions.

Research-Related Factors

When considering initial motivations for engaging with an RMT study, contributing toward novel research findings was the most prevalent reason for taking part. Across all time points, research team support was also a key facilitator of sustained engagement in the study.

Altruism and Academia

Taking part in the study was an opportunity to use personal experiences of depression to help others, to advance scientific understanding, and to “give back” to the system:

“I’ve suffered with depression the whole of my adult life, I’ve obviously had a lot out of the system. If I can do anything to put back, do you see what I mean—I will.” [P30, 24 months, United Kingdom]

Taking part for “the future, for the people who come after me” (P8, 3 months, Spain) was a strong theme that arose in all sites when discussing reasons for enrolling in the research study. Altruistic motivations continued across later time points regardless of whether participants felt they had experienced any direct benefits:

“I am actually quite proud to say that I am doing this as part of research. Some people will ask me what it is [the wearable], and I say well it is good if more people get to know about it. And for the long-term benefits, might not be for me but for other people, because it might show.” [P18, 12 months, United Kingdom]

With regard to the RMT aspect of the study, some mentioned that it “piqued my interest” (P37, 24 months, United Kingdom) and “I was very intrigued by a study that kind of has consistent monitoring” (P39, 24 months, United Kingdom). However, many participants signed up with limited knowledge of the study procedure, or of the use of RMTs for health care monitoring. Thus, a lack of prior understanding of RMTs is not a barrier to initial engagement.

Privacy was not a barrier to participants upon entering the study or throughout their participation. A key reason for this was that...
the research was conducted in a clinical and academic setting. In the Spanish cohort, one participant viewed the study as parallel to their clinical care:

> It’s not data about, about privacy, things about you, no, it’s related to a medical condition, isn’t it? A case of depression, that’s what it’s about. So if they ask you for medical data, well, it’s normal. [P25, 24 months, Spain]

Any initial privacy or data security concerns were largely alleviated by the 3-month point through conversations with the research team. At later time points, privacy was not discussed frequently.

**Research Team Support**

Support from the research team was a facilitator to continued engagement with the RMTs. This was primarily practical; at 3 months, the research team provided support on how to use the devices and study apps, which was often imperative to successful enrollment into the study:

> I tried it once [the wearable] and wasn’t able to...to...put it on the phone. If it hadn’t been for [researcher name]’s help I wouldn’t have made it. [P1, 3 months, Spain]

The need for practical support remained a key theme at 12 months, this time concerning technological malfunctions. Ability to contact the research team through various methods and receiving a timely reply was important. Some felt comfortable with initiating support themselves: “I didn’t need that much contact personally, I could get in contact easily, if it were necessary” (P21, 24 months, the Netherlands). Others wanted more contact, for example, more points of researcher-initiated contact, or specific contact from specialists. At-hand support was essential for continued participation:

> I think it is really important to have the practical support ‘cause you don’t want to be offline or not working for long than is necessary. Otherwise it goes against the purpose of the study really. [P18, 12 months, United Kingdom]

There was a consensus at all time points that the research team was approachable, patient, and reassuring, helping to alleviate technological concerns.

The research team also provided emotional support to the participants. Some participants sought comfort in the knowledge that they were being monitored as part of a study: “I liked it a lot because, jeez knowing, I felt safe, you know? Because knowing that you were there...” (P25, 24 months, Spain). Others had specific examples of receiving mental health support from the research team. One participant in the British cohort received direct signposting, which was noted in both their 12-month and 24-month interview as a crucial part of their study experience:

> because of the letter from [researcher] to the GP clinic I was able to get an immediate referral, and the problem is if you’re the system it’s great, if you’re not in the system it’s difficult to get in. I couldn’t have done it on my own. [P27, 12 months, United Kingdom]

**Benefits of RMTs for Self-management**

Despite primarily engaging with the study for altruistic reasons, many participants experienced unexpected benefits of using RMTs for symptom monitoring during their time in the study. These comprised symptom awareness and communication, both of which were integrated into self-management of depression.

**Symptom Monitoring and Awareness**

Across all 3 time points, the most frequently reported benefit was an increase in symptom awareness. Monitoring various factors related to depression, for example, mood, sleep, and exercise, increased self-reflection, and the ability to identify patterns. For example, having access to objective sleep data provided clarification and reassurance:

> I loved that [the wearable data], I found that so reassuring to just relax, of course you’ve slept and then you go ok, the next time you’re lying in bed you go I’m not ever gonna sleep again but actually you have, you’ve seen that you do I think that’s brilliant, really reassuring. [P14, 3 months, United Kingdom]

Although the app did not provide feedback on symptom scores, many felt that the act of answering the questionnaires prompted them to analyze how they had been feeling:

> I’m more aware of it, the questions on the questionnaire, especially those that ask how I’m feeling right now raise my awareness, I feel quite average or, I’m feeling not great, sometimes you ignore these things. And if you can take more time to think about these things...maybe I need to meditate more, I really feel self-conscious... [P10, 3 months, the Netherlands]

For some, answering the questionnaires and viewing the Fitbit data simply provided an understanding of their experience of depression: “I have noticed that my answers have gotten more positive throughout the year” (P22, 24 months, the Netherlands). For others, these data directly motivated behavior changes. At 3 months, the discussion focused on the motivational effects of the Fitbit data; participants felt encouraged to complete their daily step count or achieve target physical activity “badges.” Toward the later time points, these data came to act as prompts for self-care, for example, increased exercise or relaxation:

> Wearing a watch and knowing that my activity matters, you know? I mean, like the steps I take have a direct effect on my health, both physical and mental, all my activity makes me more aware of it, more conscious of it and it has also been like a driving force for me to put my batteries in sport or stress management...a habit forever, so I do not want to do without it. [P26, 24 months, Spain]

This became especially apparent during the 24-month interviews, when the Fitbit data were used to monitor sleep and mood symptom changes during the COVID-19 pandemic. Disruption to usual routines during this time allowed some to reflect more than ever on the benefit of monitoring exercise:

> I knew in theory, exercising and getting out and so on was good for your mental health, but over Covid,
the monitor helped, and the benefit would have been even better. I think I might have been worse during Covid without it. [P36, 24 months, United Kingdom]

Communication
At each time point, the RMT data were also used for communicating personal experiences to others. Participants used their increased understanding of their depression to inform others: “For the first time it kind of occurred to me to let me partner know when I could feel it was starting...so if you see my behaviour change or I’m unresponsive this is why” (P39, 24 months, United Kingdom).

Access to the Fitbit data also facilitated joint decision-making, both for immediate symptom management and long-term strategies:

There are also days that I don’t reach 5000 steps, which will make me think oh I haven’t done that many today...my spouse will say that too, go for another walk. [P2, 3 months, the Netherlands]

Overall Value and Utility
There was a consensus throughout that the benefits of participating in the study outweighed the costs, of which there were relatively few. Many had not envisioned any personal benefits when enrolling as they were aware that they would not receive personalized outcomes; however, had been pleasantly surprised by the integration of RMT data into their depression self-management, as early as the 3-month time point:

I think its empowering to know more about myself to understand more so I think once I can see more what the data is from collecting from data when the other apps are working and being able to see what the data is and notice any correlations then I think that will be really valuable. [P12, 3 months, United Kingdom]

Technology-Related Factors
Experience of the technology used in the study (smartphone apps and Fitbit) was the most widely cited theme across all sites. This covered the convenience of integrating the RMTs into daily life, the usability of the technology, technological malfunctions that occurred, and the extent to which participants found the technologies intrusive.

Convenience
Using a mobile phone and wearing a watch were already an integral part of many participants’ daily routine. The Fitbit device, “it’s basically wearing a watch” (P7, 3 months, United Kingdom), collected data passively without the need to input information, and continual wear, syncing, and charging were integrated into the routine as early as the 3-month time point. Reminder messages across the system were useful in the process of long-term integration.

One aspect that participants found more difficult to integrate into their routine was the app questionnaires. Timing of the questionnaires was often inconvenient, for example when at work, driving, or in social situations: “Obviously I’m less likely to stop my conversation to be like oh this questionnaire, because that’s a bit rude” (P4, 3 months, United Kingdom). Frequency of the ESM questionnaires was also too high from some: “it’s impossible to have a routine with that. If you have a full-time job, it’s always a bother” (P17, 24 months, the Netherlands). The participants were rarely able to change their routine to accommodate answering the questionnaires, which sometimes caused guilt. One participant in the Spanish cohort reflected on how work affected their ability to respond to app notifications during their 2-year participation:

At the beginning it was a bit difficult because I was working, then as I was on sick leave for two years, the truth is that I’ve been able to adapt quite well. And in the end, when I went back to work again, it was a bit difficult... [P1, 24 months, Spain]

Usability
For those who received a smartphone upon enrollment, a large technological barrier was the process of “relearning” a new operating system. This was described by some as “more difficult than anticipated” (P3, 3 months, United Kingdom), particularly during the 3-month interviews, owing to adapting to a new user interface and decreased connectivity with other devices. At 24 months, some participants had adjusted to using the new device, whereas others planned to swap back upon study completion:

No, my only peeve was that I’m an Apple user and having this bloody awful Android phone, the first thing I shall do on April 1st is take my SIM card out of the Motorola thingy. [P35, 24 months, United Kingdom]

Technological Malfunctions
The participants reported a range of technological malfunctions that affected their participation in the study. Issues with the study apps were particularly prevalent during the 3-month interviews owing to ongoing technological challenges during the early phases of the study. These included not receiving notifications, apps crashing, apps logging out, and difficulties with rescanning QR codes. Participants sometimes had limited time or motivation to report issues to the team:

I tried opening a questionnaire I wouldn’t be able to see it, I wouldn’t be able to do it and there was no way of saying this is happening or why this is happening so maybe I should have contacted you about it but I just kind of ignored it. [P4, 3 months, United Kingdom]

Issues with missing data persisted throughout the 3 time points. Participants were aware of the times when the active app had been unable to submit the completed data, or the passive app had ceased monitoring. Such malfunctions often led to anxiety or guilt that they were not “correctly” participating: “Well, yes, when it didn’t work, I became a bit nervous...” (P15, 3 months, Spain).

Participants also reported frequent missing data with the Fitbit, caused either by a syncing error or inaccurate recording. These issues caused some to question the integrity of the study: “It just didn’t work and that’s not what you expect from a research study” (P18, 24 months, the Netherlands).
A participant in the Spanish cohort reflected on how these technological malfunctions affected not only their ability to participate in the study but also their experience of being able to use the resulting data:

> There is data that I have missed here, and of course I was analyzing it with me in important situations of how I was, and that I have missed them, for more than a month. [P32, 24 months, Spain]

**Intrusiveness**

Generally, the concept of remote monitoring, or the use of the technologies, was not regarded as intrusive. Rather, passive data collection was noted as a preferable method because “at some point you don’t notice it. You don’t notice that you’re wearing it anymore” (P18, 24 months, the Netherlands).

However, one area that caused disruption was the wearability of the Fitbit device. Several issues associated with the Fitbit strap were reported, including skin irritation, increased sweating, and allergic reactions. Some had briefly chosen to remove the device while experiencing discomfort, whereas others had purchased straps with alternative materials. At 12 months, many reported that their strap had broken, and by 24 months, some had to apply for a full device replacement. One participant felt guilty when asking the research team for their device to be repaired:

> I know that the money allocated to research programs or projects is minimal, and of course, when the strap broke or the Fitbit wouldn’t charge me and then I felt really bad because I thought “oh my God, now they have to change my Fitbit.” [P26, 24 months, Spain]

Waiting for a replacement strap or device meant that participants were unable to continue to use the Fitbit for self-management:

> if I was going to continue and for the others who will be continuing, it will probably begin to happen more and more depending on how much people are actually exercising with them on. It only grows, that’s the problem, in my experience with the other Fitbit, that definitely happens. [P3, 12 months, United Kingdom]

**Clinical Factors**

The participants were asked to reflect on whether and how they expected RMT data to be beneficial in clinical care. These included (1) allowing the clinician to view the “whole picture” of individual experience, (2) allowing the clinician insight into new symptoms, (3) as a way for patients to report specific areas of concern, and finally (4) as a basis for making decisions about suitable treatment or care. Importantly, treatment decisions should be reached as a joint decision involving the clinician, the patient, and the data:

> I think they could actually look at the data that’s being produced, and that could assist them in helping me to come to another decision. Like, if I was deciding that I would like to move my medication down, but they’ve got the data that says, no you’re not...but if it backs it up as well, so it can work both ways, so I think it does have those benefits. [P33, 24 months, United Kingdom]

Sleep data were repeatedly cited as a data stream that would cause change in treatment. Participants from all sites provided examples of conversations with their mental health clinicians. One participant in the British cohort also discussed their experience of integrating the sleep data into their sleep clinic appointments:

> It’s too expensive for the NHS to keep on doing [sleep tests]...I said, well, actually, I can show you any time in the last six months or so...an indication of when I’m sleeping...It helped them choose what exercises I needed to do and what therapy was required, so, yes, it was extremely helpful. [P22, 12 months, United Kingdom]

Presentation of objective sleep data was seen as helpful “proof” of the participant’s recent experiences:

> You can tell your GP that you sleep terribly, but of course your GP can also think that you’re just worried, but with the data it’s a fact that you can prove, so that’s nice, that you have concrete info...whether you worry or complain about it or not doesn’t matter, the facts are there. [P10, 12 months, the Netherlands]

**Current Clinical Utility of RMTs**

Although the potential for RMTs in clinical care was recognized, 2 key barriers to their implementation were envisioned. First, the level of technological acceptance of medical professionals...
influenced participant views on the long-term utility of the data. Participants in the Spanish cohort, who were recruited through their clinical care, generally reported acceptance of the study from their clinicians: “even my psychiatrist here and in Barcelona had the same way of thinking and saw that this was very useful for me and encouraged me” (P9, 24 months, Spain).

Others described more negative experiences, often causing them to question the use of the data:

I thought it would be more relevant for my neurologist, but my neurologist wasn’t particularly interested when I told him about what I was doing in the study. [P17, 12 months, United Kingdom]  
Second, lack of funding, resources, and time was perceived as a major roadblock to using RMT data in appointments. This was particularly apparent in the British cohort with regard to the National Health Service. For the data to be monitored and reflected on, new procedures would need to be put in place:  
I would be amazed if there was sufficient funding for that...I don’t believe that the NHS have got the resources to have people monitoring this sort of stuff. [P22, 12 months, United Kingdom]  
Given the perceived lack of resources to effectively use RMT data in the National Health Service, some have considered how best to come to a compromise:  
I think realistically, if they had that [data] and I went to them with a problem, then I would like them to be able to use it at that point. But I don’t see it as something that they would be—so, for example, if I went to them with something and if somehow, it was a part of my NHS records, if they could access that, that might be helpful to them. But I don’t see them using it other than that really. [P32, 24 months, United Kingdom]  

System Amendments and Additions
Participants discussed various changes or additions to the RMT system used in this study to further encourage long-term engagement. These included suggestions for questionnaire data collection and feedback.

Data Collection
Across all sites and time points, the most prevalent suggestions for changes to the study design were the content of active RMT questionnaires. Participants felt that they were frequently being asked to complete the same questions, particularly within the ESM schedule, which often prompted them to provide the same answers, for example, with regard to mood changes. This affected motivation:

At first, I was more excited about it, but as time has passed, sometimes I don’t feel much like answering since the same questions get repeated. [P19, 12 months, Spain]  
Some also suggested the ability to postpone questionnaires if feeling too low to complete them and the ability to provide contextual information. As early as the 3-month time point, some noted that external factors affecting their mood were not being monitored within the validated mood and self-esteem questionnaires: “I notice that when my home situation isn’t great, I also fill in the questionnaires less positively” (P5, 3 months, the Netherlands). On reflection, some would have liked to have given more information at certain points:

The answers are very closed, so you can’t really answer what you feel. You know? It’s very...it’s very up in the air. [P1, 24 months, Spain]  

Data Feedback
When asked how they might wish to view their symptom data in future use, the majority felt that this was best displayed visually through in-app graphs. Many also expressed that this would need to be accompanied by a “human explanation for what those things mean” (P3, 12 months, United Kingdom). There was a discrepancy between when these data would be best received; some only expected to receive it at the end of the study, some felt that it would be more useful in real time, whereas others were cautious that receiving data during periods of low mood would be detrimental:

If I’m well I want to see it, if I’m unwell, no. If I was reporting that I was feeling suicidal I don’t think I’d want to revisit it. [P27, 24 months, United Kingdom]  
Furthermore, some participants considered the potential for RMT data to provide feedback on symptom patterns and changes over time, correlations with other factors, and depressive relapse prediction. Specific examples included relationships between exercise and mood, sleep and mood, and mood and concentration: “At some point I had a burn out. I’m very curious as to how my ability to concentrate changed, and if that maybe shows on the THINC-it app” (P3, 24 months, the Netherlands).

It was generally accepted that having access to data of this nature would be useful for both self-management and integration into clinical care. Looking forward at the 24-month time point, one participant at the British site explained their hopes for the future of this field:

I think trends are really quite important for me in managing what is going on...I think one of the things I am thinking would be good to come out of this is an ability to see patterns over time and then maybe being able to use that as a predictor or, I need to do some intervention here so that I don’t end up there again if that makes sense. [P30, 24 months, United Kingdom]  

Discussion

Principal Findings
An exploration of the subjective experience of long-term engagement with RMTs for depression symptom management could prove a necessary complement to objective engagement statistics, providing insights into technology usability, user experience, and facilitators of sustained use. This study aimed to (1) explore the key themes associated with long-term RMT use and (2) identify recommendations for future engagement through longitudinal qualitative analysis at 3-month, 12-month, and 24-month time points of the RADAR-MDD study.
The themes uncovered suggest that long-term engagement with RMTs can be understood from two main perspectives: (1) experiential factors and (2) system-related factors (Figure 1). Experiential factors relate to the ways in which participants construct their experiences of engaging with RMTs for symptom monitoring. Experiential factors comprise research altruism, support from a professional team, and the benefits of using RMTs for depression management. System-related factors refer to direct engagement with the RMT systems. The factors include the usability, convenience, and intrusiveness of the technologies and the recommended system improvements for successful clinical implementation.

On the basis of these perspectives, we present a set of considerations for the promotion of engagement with RMTs for depression. Given the breadth of use cases proposed for RMTs in MDD, we focused on two areas: (1) engagement with research and (2) engagement with real-world implementation. Recommendations for engagement with future RMT research are outlined in Multimedia Appendix 3.

Although our data were derived from research participants, we believe that our findings can also be useful when considering implementation into clinical practice. Participants identified the following opportunities for RMTs in clinical care: (1) provision of feedback-informed care, (2) strengthening the therapeutic relationship, and (3) the specific clinical value of sleep monitoring. However, this potential was acknowledged with the caveat of a perceived lack of time and resources in clinical care across all 3 countries. Our findings indicate that a large difference between engagement with RMTs for research and long-term clinical engagement could be research altruism. In this study, an important facilitator of both initial and sustained engagement was the experiential factor of taking part in a novel, academic study to advance understanding and help others. To this end, participants forewent privacy concerns and initial receipt of personal benefit. They were also willing to engage despite the implementation concerns. In the absence of research altruism, Figure 1 can be used to identify further experiential facilitators that could instead be harnessed to promote engagement when RMTs become integrated into evidence-based practice. For example, clinical onboarding sessions could include a clear summary of the proposed uses and benefits of RMT data and symptom monitoring for an individual’s care. Multimedia Appendix 4 provides a set of considerations for the implementation of RMTs into clinical care based on the experiential and system-related factors identified.

Figure 1. Experiential and system-related factors in the subjective experience of longitudinal remote measurement technology (RMT) use.
Comparison With Previous Work

This study builds on previous qualitative analyses of the barriers to and facilitators of intended RMT use for depression management. The functional and nonfunctional requirements set out by Simblett et al. [22] roughly align with the system and experiential factors found here. However, a comparison of coding frameworks (Multimedia Appendix 2) revealed several differences in this study. First, nonfunctional, user-related factors such as cognition, symptom severity, and emotional resources were not acknowledged as barriers to long-term RMT engagement. Second, the overall utility of RMTs was discussed mainly in terms of benefits and rewards, and less so in terms of costs such as privacy and security. Third, studying long-term RMT use has revealed an additional layer of understanding surrounding nonfunctional requirements; experiential factors include the impact of professional support and the effects of symptom monitoring on self-awareness and communication.

When comparing our findings with those from the wider mHealth literature, technological and system-related factors remained a common theme. Borghouts et al. [26] and Patel et al. [27] found that lack of technical issues, flexible usability of the platform, personalization, and access to training were associated with increased long-term engagement with digital health intervention platforms. One clear difference with digital health intervention work is the focus on “a desire to actively improve one’s health” [27] as a main facilitator of initial and sustained engagement. Our work has shown that in the absence of a direct or tangible benefit, users remain willing to interact with RMTs for long periods within a research context.
Experiential factors such as advancing scientific understanding and, at later periods, experiencing indirect benefits of mood tracking, seem to operate as a supplement to the user-related factors currently reported in the field.

**Strengths and Limitations**

To the best of our knowledge, this is the largest study to qualitatively explore long-term RMT use for depression across multiple countries. Data collection and analyses were conducted in the native language of each country and only quotes were translated into English, aiding the transfer of meaning process [34]. However, this study has some limitations. First, where we did not anticipate any major intercountry differences in terms of attitudes toward remote mental health tracking, participants in the Spanish cohort were invited to participate by the clinicians involved in their care. This might have overinflated some themes in our analyses; for example, perceived benefits of the technologies. Second, interviews were conducted via convenience sampling of the participants who remained enrolled at each time point. This increased the risk of selection bias; those who enjoyed using the RMTs were more likely to continue to engage and as a result more likely to agree to an interview. This could explain the absence of themes relating to symptom severity or cognitive barriers present in the current work, although recent analyses have suggested that these factors did not contribute to sustained engagement in the study [35]. Convenience sampling also resulted in 21 participants completing the interviews at ≥2 time points. Preliminary sensitivity checks on a subset of this sample showed no clear signs of changes in themes over time. The data were not deemed rich enough to undertake a full, longitudinal analysis on this sample. Third, because of resource constraints, no sites undertook double coding. Fourth, data-driven themes were not explored in relation to demographic or clinical factors, as this was deemed beyond the scope of this study. Although previous work suggests that perceived usability, and actual use, of the RADAR-base system remains robust across severity of clinical characteristics [35], understanding demographic differences in subjective engagement is an important avenue for future research. Finally, the COVID-19 pandemic occurred during the study follow-up period. Given the transition to remote working and health care across all 3 countries during this time, the subjective experience of using RMTs might have been positively skewed; for example, with regard to the positive impact of the research team during social isolation. It should also be noted that the topic guide primarily asked participants to review their experience of using RMTs for this specific research project, and specific use cases for clinical implementation were not outlined by interviewers. Thus, the themes that arose from this work relate primarily to long-term engagement with RMT research, and the transferability of the findings to engagement in clinical care should be taken with caution.

**Applications for Future Research**

Future work should continue to explore subjective engagement with RMTs, conceptualized in terms of both experiential and system-related factors. Where system-related factors often represent clear recommendations for technological improvements, understanding the experiential effects of engaging with RMTs is a novel finding that could prove fundamental in promoting future engagement. A recent systematic review [17] found that 5 studies have begun to explore the correlational relationship between objective and subjective engagement with RMTs. Higher daily assessment counts from an active RMT app were correlated with increased app satisfaction ratings at 3-month and 6-month time points [36,37]. Understanding the link between experiential factors, such as increased self-awareness, and objective engagement could bolster this field further.

Our findings explore the initial and sustained engagement with RMTs for depression symptom monitoring in a research setting. The next step would be to replicate this work in a clinical setting. Recent qualitative analyses have reported positive views from patients and clinicians on the potential for implementation of RMT into psychological services [38]. This paper provides considerations for adapting RMT systems for use in clinical settings and a framework for continuing to analyze the subjective experience of long-term clinical engagement to allow for further iterations.

**Conclusions**

This study aimed to understand the subjective experience of long-term engagement with RMTs for depression symptom monitoring as a complement to the high rates of objective engagement observed in the RADAR-MDD study. Key experiential and system-related themes associated with long-term RMT use were identified along with a set of recommendations and considerations for promoting future system use in both research and clinical settings. Further understanding of the construction of the “experience” of using RMTs will be key to promoting long-term engagement in clinical care and depression management in comparison with general mHealth interventions that offer immediate or tangible rewards. In the wake of the rapid expansion of this field, we urge professionals to continue monitoring the subjective experience of RMT engagement to maximize the potential of remote monitoring as both a method for data collection and a tool for symptom management.

**Acknowledgments**

This paper represents an independent research part funded by the National Institute for Health Research (NIHR) Maudsley Biomedical Research Centre at South London and Maudsley National Health Service (NHS) Foundation Trust and King’s College London. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care.

The authors would like to thank all the members of the Remote Assessment of Disease and Relapse-Central Nervous System (RADAR-CNS) patient advisory board for their contribution to the device selection procedures and their invaluable advice.
throughout the study protocol design. The authors would also like to acknowledge the work of Felice Fernhout in conducting coding on the data.

Participant recruitment in Amsterdam was partially accomplished through Hersenonderzoek.nl, a Dutch web-based registry that facilitates participant recruitment for neuroscience studies [39]. Hersenonderzoek.nl is funded by ZonMw-Memorabel (project number 73305095003), a project in the context of the Dutch Deltaplan Dementie, Gieskes-Strijbis Foundation, the Alzheimer’s Society in the Netherlands, and Brain Foundation Netherlands. Participants in Spain were recruited through the following institutions: Parc Sanitari Sant Joan de Déu network of mental health services (Barcelona), Institut Catalá de la Salut primary care services (Barcelona), Institut Pere Mata-Mental Health Care (Terrassa), and Hospital Clínico San Carlos (Madrid).

The authors would like to thank all Genetic Links to Anxiety and Depression study volunteers for their participation and gratefully acknowledge the NIHR BioResource, NIHR BioResource centers, NHS Trusts, and staff for their contributions. The authors would also like to acknowledge NIHR Biomedical Research Centre (BRC), King’s College London, South London and Maudsley NHS Trust and King’s Health Partners. The authors would like to thank the NIHR, NHS Blood and Transplant, and Health Data Research United Kingdom, as part of the Digital Innovation Hub Program.

This research was reviewed by a team with experience of mental health problems and their caregivers, who were specially trained to advise on research proposals and documentation through the Feasibility and Acceptability Support Team for Researchers (FAST-R): a free, confidential service in England provided by the NIHR Maudsley BRC via King’s College London and South London and Maudsley NHS Foundation Trust.

The RADAR-CNS project received funding from the Innovative Medicines Initiative (IMI) 2 Joint Undertaking under grant 115902. This joint undertaking received support from the European Union’s Horizon 2020 research and innovation programme and European Federation of Pharmaceutical Industries and Associations. This communication reflects the views of the RADAR-CNS consortium and neither IMI nor the European Union and European Federation of Pharmaceutical Industries and Associations are liable for any use that may be made of the information contained herein. The funding body was not involved in study design, data collection or analysis, or data interpretation.

Data Availability
The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Authors’ Contributions
KMW contributed to the design and coordination of the study in London as well as to the data processing, coding, analysis, and writing of this manuscript. EDL contributed to the data coding and analysis. S Siddi contributed to the design and coordination of the study in Barcelona as well as the data coding. FL contributed to the design and coordination of the study in Amsterdam, as well as the data coding. S Simblett contributed to the development and design of the study and advised on the analyses. GRA has contributed to data coding. AI contributed to the study conducted in London. IM-G contributed to the development and design of the study. JMH contributed to the development and design of the study. CO contributed to the study conducted in London. PP contributed to the data coding and analysis. AR contributed to the development and design of the study. ER contributed to participant recruitment for the study. TW contributed to the development and design of the study. CH contributed to data interpretation and supervision of the first author. MH secured funding and is the principal investigator of the study, and contributed to the overall study design and conduct. FM contributed to the design and coordination of the study. Patient advisory board members contributed to the design and development of the study.

Conflicts of Interest
MH is the principal investigator of the RADAR-CNS program, a precompetitive public-private partnership funded by the Innovative Medicines Initiative and the European Federation of Pharmaceutical Industries and Associations. The program received support from Janssen, Biogen, Merck & Co, Union Chimique Belge, and Lundbeck. JMH has received economic compensation for participating in advisory boards or giving educational lectures from Eli Lilly & Co, Sanofi, Lundbeck, and Otsuka. CO is supported by the UK Medical Research Council (MR/N013700/1) and King’s College London member of the MRC Doctoral Training Partnership in Biomedical Sciences.

Multimedia Appendix 1
Main interview questions at 3-month, 12-month, and 24-month follow-up time points.
[DOCX File, 21 KB - humanfactors_v10i1e39479_app1.docx ]

Multimedia Appendix 2
Preliminary and final codes in the coding framework.
[DOCX File, 20 KB - humanfactors_v10i1e39479_app2.docx ]
Multimedia Appendix 3
Recommendations for future remote measurement technology (RMT) use in observational research.
[DOCX File, 319 KB - humanfactors_v10i1e39479_app3.docx]

Multimedia Appendix 4
Considerations for remote measurement technology (RMT) implementation in real-world clinical settings.
[DOCX File, 366 KB - humanfactors_v10i1e39479_app4.docx]

References


38. de Angel V, Lewis S, White KM, Matcham F, Hotopf M. Clinical targets and attitudes toward implementing digital health tools for remote measurement in treatment for depression: focus groups with patients and clinicians. JMIR Ment Health 2022 Aug 15;9(8):e38934 [FREE Full text] [doi: 10.2196/38934] [Medline: 35969448]

39. Hersenziekten de wereld uit helpen kan alleen met onderzoek. hersenonderzoek nl. URL: https://hersenonderzoek.nl/ [accessed 2023-01-13]

Abbreviations

ESM: experience sampling methodology
MDD: major depressive disorder
mHealth: mobile health
RADAR-MDD: Remote Assessment of Disease and Relapse-Major Depressive Disorder
RMT: remote measurement technology
Pediatric Primary Care Providers’ Perspectives on Telehealth Platforms to Support Care for Transgender and Gender-Diverse Youths: Exploratory Qualitative Study

Gina M Sequeira¹,², MD, MS; Nicole F Kahn¹, MEd, PhD; Kevin M Bocek¹, BA; Taraneh Shafii², MD; Peter G Asante², MD; Dimitri A Christakis¹,², MPH, MD; Wanda Pratt¹, PhD; Laura P Richardson¹,², MPH, MD

¹Seattle Children’s Research Institute, Seattle, WA, United States
²Department of Pediatrics, University of Washington, Seattle, WA, United States
³Information School, University of Washington, Seattle, WA, United States

Corresponding Author:
Gina M Sequeira, MD, MS
Seattle Children’s Research Institute
4540 Sand Point Way NE Building 1
Seattle, WA, 98105
United States
Phone: 1 206 987 2028
Email: gina.sequeira@seattlechildrens.org

Abstract

Background: Access to gender-affirming care services for transgender and gender-diverse youths is limited, in part because this care is currently provided primarily by specialists. Telehealth platforms that enable primary care providers (PCPs) to receive education from and consult specialists may help improve the access to such services. However, little is known about PCPs’ preferences regarding receiving this support.

Objective: This study aimed to explore pediatric PCPs’ perspectives regarding optimal ways to provide telehealth-based support to facilitate gender-affirming care provision in the primary care setting.

Methods: PCPs who had previously requested support from the Seattle Children’s Gender Clinic were recruited to participate in semistructured, 1-hour web-based interviews. Overall, 3 specialist-to-PCP telehealth modalities (tele-education, electronic consultation, and telephonic consultation) were described, and the participants were invited to share their perspectives on the benefits and drawbacks of each modality, which modality would be the most effective, and the most important characteristics or outcomes of a successful platform. Interviews were transcribed and analyzed using a reflexive thematic analysis framework.

Results: The interviews were completed with 15 pediatric PCPs. The benefits of the tele-education platform were developing a network with other PCPs to facilitate shared learning, receiving comprehensive didactic and case-based education, having scheduled education sessions, and increasing provider confidence. The drawbacks were requiring a substantial time commitment and not allowing for real-time, patient-specific consultation. The benefits of the electronic consultation platform were convenient and efficient communication, documentation in the electronic health record, the ability to bill for provider time, and sufficient time to synthesize information. The drawbacks of this platform were electronic health record–related difficulties, text-based communication challenges, inability to receive an answer in real time, forced conversations with patients about billing, and limitations for providers who lack baseline knowledge. With respect to telephonic consultation, the benefits were having a dialogue with a specialist, receiving compensation for PCP’s time, and helping with high acuity or complex cases. The drawbacks were challenges associated with using the phone for communication, the limited expertise of the responding providers, and the lack of utility for nonemergency issues. Regarding the most effective platform, the responses were mixed, with 27% (4/15) preferring the electronic consultation, 27% (4/15) preferring tele-education, 20% (3/15) preferring telephonic consultation, and the remaining 27% (4/15) suggesting a hybrid of the 3 models.

Conclusions: A diverse suite of telehealth-based training and consultation services must be developed to meet the needs of PCPs with different levels of experience and training in gender-affirming care. Beyond the widely used telephonic consultation model, electronic consultation and tele-education may provide important alternative training and consultation opportunities to facilitate greater PCP independence and promote wider access to gender-affirming care.
Introduction
Background
As the population of youths who identify as transgender and gender-diverse (TGD) youths continues to grow [1], the need for gender-affirming care in pediatric settings substantially exceeds availability, leaving many TGD youths who are interested in receiving this care without access to it [2]. Given that the existing research suggests that access to gender-affirming care during adolescence is associated with improved mental health outcomes [3-8], increasing the availability of this care for TGD youths is critical. Currently, the provision of gender-affirming care is largely limited to specialty clinics located within pediatric hospital systems in large urban areas [9-15]. One way to improve the access and remove the barriers to gender-affirming care is by providing such care in the primary care setting. However, only a few pediatric primary care providers (PCPs) have received training in gender-affirming care [13-15], and many are unaware of how to create affirming environments and discuss treatment options available to TGD youths. Therefore, pediatric PCPs need opportunities to receive education from and consult gender specialists.

Prior Work
Telehealth has the potential to meet these needs and is an umbrella term that describes both patient-to-provider audio-video visits (telemedicine) and specialist-to-PCP consultation methods, such as tele-education, electronic consultation, and telephonic consultation. Tele-education platforms connect cohorts of PCPs with specialists via the web for live didactic education and case consultation sessions. This modality, which has been used to facilitate gender-affirming care provision to TGD adults [16,17], increases provider knowledge and improves PCP’s clinical confidence [18]. Electronic consultation uses store-and-forward electronic dialogue to provide patient-specific, specialist-to-PCP consultation. This modality, which has also shown great promise in facilitating gender-affirming care provision in primary care for TGD adults [19-22], has led to increased provider knowledge along with decreased barriers to accessing specialty care [23-25]. Finally, telephonic consultation, which is the most common of these consultation models, typically involves PCPs calling an on-call specialist to discuss a specific case over the telephone [26]. These informal consultations, often referred to as “curbside” consultations, have raised concerns among specialists regarding the quality of care, patient safety, documentation, and compensation [26,27].

Goal of This Study
Given the increasing prevalence of gender diversity [1], the inaccessibility of pediatric gender-affirming care among many youths [2,12], and the lack of training among pediatric PCPs, we must develop specialist-to-PCP telehealth platforms to guide PCPs in providing gender-affirming care. These platforms are critical because they can provide remote training and consultation, thus broadening the reach of pediatric gender-affirming care services to diverse and underresourced settings and populations. To our knowledge, no prior studies have been conducted with pediatric PCPs about how best to use specialist-to-PCP telehealth platforms, such as tele-education, telephonic consultation, and electronic consultation, to support them in providing gender-affirming care to TGD youth. Thus, the purpose of this qualitative study was to explore pediatric PCPs’ perspectives regarding optimal ways to provide telehealth-based support to facilitate gender-affirming care provision in pediatric primary care settings.

Methods
Recruitment
Potential participants were identified from a list of community pediatric providers across Washington, Wyoming, Alaska, Montana, and Idaho who had either previously (1) called or emailed the Seattle Children’s Gender Clinic (SCGC) team for support with a patient management question or (2) attended continuing medical education training provided by SCGC in the 2 years before recruitment. The participants were recruited via email by a member of the research team who provided an overview of the proposed study. Invitations to participate in the study were sent to 69 potential participants via email. Of these 69 individuals, 20 (33%) completed a screening survey that was used to determine study eligibility. Eligibility criteria included the following: (1) currently in practice delivering ambulatory primary care to patients aged <18 years, and (2) able to complete an hour-long audio-video interview via Zoom (Zoom Video Communications, Inc) [28]. Of 20 individuals who completed the survey, 15 agreed to participate in a semistructured interview with a member of the research team trained in qualitative research.

Data Collection
Demographic information was collected from the screening survey and included participants’ age, race, ethnicity, gender identity, years in practice, and practice location (urban, rural, or suburban); the number of TGD youths they have seen in their practice; and the number of patients they have referred to SCGC. Details regarding tele-education preferences were collected using semistructured interviews based on guides developed with input from 3 pediatric PCP stakeholders and a community advisory board of TGD youths and their parents. The interview guide consisted of 2 parts. The first part explored PCPs’ perspectives regarding their role in providing gender-affirming care and the barriers they have faced in the primary care setting. The second part of the interview presented 3 different telehealth modalities (tele-education, electronic consultation, and telephonic consultation) using standardized definitions (Multimedia Appendix 1), and each participant was asked to
share their perspectives on the following: (1) the benefits and drawbacks of each modality, (2) which modality would be most effective in supporting them in providing gender-affirming care in the primary care setting, and (3) the most important characteristics or outcomes of a successful platform. The data presented in this paper are limited to those collected in the second portion of the interview.

**Data Analysis**

Interview transcripts were automatically generated via Zoom with an embedded transcription software and were cleaned and corrected by 2 trained research coordinators. Then the transcripts were independently coded by 2 members of the research team using a codebook consisting of 64 codes that was developed in partnership with a PCP stakeholder who is currently providing gender-affirming care to TGD youths. Themes were then iteratively generated using a reflexive thematic analysis framework [29]. Coding was performed using the qualitative analysis software Dedoose (Socio Cultural Research Consultants, LLC) [30].

**Ethics Approval**

The participants provided informed consent to participate in the study and received a gift card worth US $20 for their participation. All study procedures were approved by the SCGC’s institutional review board (STUDY00002986) before recruitment.

**Results**

**Participant Characteristics**

Interviews were completed with 15 pediatric medical providers, including advanced practice providers, pediatricians, and family medicine physicians currently providing outpatient clinical care to youths aged <18 years. The participants represented a wide range of years in practice, with one-third (5/15, 33%) having practiced for >10 years and 40% (6/15) having practiced for <5 years. Nearly half (7/15, 47%) of the participants practiced in an urban area, whereas the remaining practiced in rural (4/15, 27%) and suburban (4/15, 27%) environments. Roughly half (7/15, 47%) of the participants indicated that they had seen >15 TGD patients, whereas one-third (5/15, 33%) indicated that they had seen ≤10 TGD patients. Finally, 40% (6/15) of the participants indicated that they had referred >5 patients to a gender clinic for care, with the remaining 60% (9/15) stating that they had referred ≤5 patients.

**Perspectives on the Proposed Telehealth Platforms**

**Tele-Education Platform**

With respect to the advantages of the tele-education platform (Textbox 1), the following themes were identified: (1) developing a network with other PCPs that facilitates shared learning; (2) receiving comprehensive, didactic, and case-based learning; (3) having scheduled education sessions; and (4) increasing provider confidence in delivering gender-affirming care.
**Textbox 1.** Themes primary care providers (PCPs) identified as benefits and drawbacks of the tele-education platform.

### Benefits
- **Theme 1: developing a network with other PCPs facilitates shared learning**
  - “You can develop some sense of...community and get to know other providers who are doing similar work nearby.”
  - “It brings communities together. So, it breaks down the siloed walls of different institutions where we can really support each other...”
- **Theme 2: receiving comprehensive, didactic, and case-based learning**
  - “I think the benefit being that you can get...more in-depth, education, you can get a good overview rather than you saying ‘Gosh I think I need to know more about this,’ you know somebody else can say ‘You need to know about this, and this, and this, and this,’ because I’m probably going to be missing something if I just pick it up myself...So...a more overarching education, probably be able to get more in depth...”
  - “I think...even if I didn’t have a case to bring, I think I could learn. Or...if I have a patient that’s similar to that I could...learn from that and potentially implement something”
- **Theme 3: education sessions occur at a scheduled time**
  - “Another thing would be if it’s predictable...if it’s once a month, at the same time, it’s something where people could plan their schedule around it and just have it already known that they’re going to talk about, they’re going to be available that time and they can make it.”
- **Theme 4: increases provider confidence in delivering gender-affirming care**
  - “I think [tele-education] will break down the fear of starting gender affirming healthcare for a lot of people out there, especially if they’re able to kind of walk through things.”

### Drawbacks
- **Theme 1: requiring a significant time commitment**
  1a: hard to prioritize over other training opportunities
  - “It’s hard to prioritize as a clinician. They’re often like...12 sessions...over a course of three months. And there’s a lot of requests for different ECHO trainings...so I tend to do one per year and I’m not sure if I would choose to do one on...transgender care now, but maybe historically would have.”
  1b: may not be worth the time investment for PCPs who see fewer transgender and gender-diverse patients or those in close proximity to a gender clinic
  - “I don’t know if I have enough volume in my clinic to have up to date, questions, or case studies. It’s a low volume, kind of, high acuity thing.”
  - “I think the cynical side would be in New Mexico, if you’re practicing in a rural place where...it’s just you, you are motivated to...fix that liver failure. But...when you’re practicing in [a large urban area with a gender clinic] and you’re like...I could...really invest a lot of time into doing this, or I could...just write a referral.”
  1c: would take time away from clinical care and decrease productivity
  - “It does seem like it’s more resource intense because obviously you have to take time out from your clinical practice, and you have to have support of your supervisors and there’s probably some financial impact to that, and it takes more time to get to the end result.”
  1d: difficult to schedule at a time that is convenient
  - “I think one of the drawbacks is finding a time where everybody can access it...we can’t have this three days a week, every month, just so everybody can access it. So, I think that can be a little bit difficult....scheduling...puts a lot of burden on the people who hold the [tele-education] program.”
- **Theme 2: not allowing for real-time, patient-specific consultation**
  - “I think the main drawback would be timeliness...I would have somebody in my office today, and would have a question about treatment or something of that sort, and I would have to wait two weeks, and remember to get back with them to tell them what to do and perhaps the thing that might be needed would need to be done fairly quickly.”
  - “When you do case presentations for patients there are supposed to be no identifiers. Oftentimes I really need to talk about this specific patient and what’s going on.”

Regarding networking, PCPs appreciated getting to know their peers who were doing similar work and developing relationships for future collaboration. They also felt that having a didactic component and listening to other PCPs’ case presentations were
important to increase their knowledge about gender-affirming care provision:

I think one of the main benefits is feeling connected to other providers in your community or beyond your community...there can be a lot of isolation in primary care when you’re providing services that aren’t provided by everybody. So, I think that’s really awesome, the community aspect of it. And I think that hearing other people talk about their cases is really valuable...listening to my colleagues present is always something that’s really interesting to me and I feel like I learned a lot that way.

In addition, the PCPs appreciated that the tele-education sessions typically took place at a scheduled time, making it easier for them to coordinate with their clinical schedules.

The participants, particularly those who already had some training in gender-affirming care, felt that having access to tele-education would help increase their confidence in providing gender-affirming care, especially during the early stages of providing such care:

Especially in a time when you’re doing...information gathering to see if it’s something that is transferable to your clinic environment, [tele-education] can be really valuable...honestly if something like this in the beginning had existed, I would have been very likely to take it on.

In terms of the drawbacks of the tele-education platform, the PCP-identified themes were as follows: (1) requiring a significant time commitment and (2) not allowing for real-time, patient-specific consultation (Textbox 1). Regarding time, some PCPs noted not feeling that they would be able to commit enough time to participate. This was especially true for providers who reported seeing a fewer number of TGD patients and those practicing in close proximity to a gender clinic. PCPs also cited concerns that participating in the tele-education platform would decrease their clinical productivity:

In a system where we are paid on productivity...me taking two hours to go to a tele-education thing is six patients that I’m not seeing, right? Which is...25% of my patient load for the day, which is 25% less pay. Right? And...it’s not really about the money, but...I’m held to a productivity standard. If I’m not meeting that...I think you would lose people. Because you either have to do it before or after work or they have to do it instead of seeing patients.

PCPs also mentioned feeling that it would be very difficult to schedule the sessions at a time that is mutually convenient for a large group of providers. Finally, PCPs noted that in comparison with electronic and telephonic consultation, the cadence with which scheduled tele-education sessions take place would limit their ability to receive support regarding patient-specific management questions:

[The tele-education session is] probably happening once a month or every other month, so if you had a case and it just happened last week, you’re now waiting two months to present this patient.

Electronic Consultation

With respect to the advantages of the electronic consultation model, themes were identified: (1) convenient and efficient communication, (2) documentation in the electronic health record (EHR), (3) ability to bill for provider time, and (4) sufficient time to synthesize information (Textbox 2).
Textbox 2. Themes primary care providers identified as the benefits and drawbacks of electronic consultation.

Benefits

- Theme 1: convenient and efficient communication
  - “We have patients constantly throughout the day, but we have a few minutes here and there where we can, finish up typing and talk to this person. So having electronic consultation would be amazing because I could just quickly type in my question. And then knowing that no one’s expecting an immediate response, I could go back and see some patients and could carry on with my day and then, when the consultation comes back in I can use my few minutes between the next patient and look at it. We as primary care providers seem to have like three to five minutes here and there throughout the day. We don’t have a full twenty minutes or half an hour to be on the phone conducting [a telephonic consultation]”

- Theme 2: documentation in the electronic health record (EHR)
  - “So I think having the electronic record to be able to refer back to would be awesome. Because maybe the question you asked about for one patient will apply to a patient in the future, so, then you can just reference back to it, I think that’s a huge strength”

- Theme 3: ability to bill for provider time
  - “We have this like psychiatrist who works with us now that I can actually e-consult, which is great...Because he needs the time for this and he’s consulting for all of us, it will be an official consult that’s billed to insurance.”

- Theme 4: sufficient time to synthesize information
  - “I would probably be more likely to use a web base or electronic consultation, because sometimes you just don’t have time in clinic to say everything you need to say. And sometimes you, as a medical provider, need to like, sit down and think about it to be like ‘What is my question?’”

Drawbacks

- Theme 1: EHR-related difficulties
  - 1a: incompatibility of EHR with specialists
    - “You know, so you’ve got some people that are like on a Cerner platforms, some are on Epic, some on all scripts, some are next gen. You know, we’re still not in this place where we have standardized the utility of our electronic health records and they don’t talk to each other, so I think that that could be problematic.”

  - 1b: using an unfamiliar EHR to submit clinical questions
    - “When we do use Epic with the other clinics system...they’re always, like, ‘Where are the labs?’ And then you say I ‘I sent the labs and here they are again.’ And they’re like, ‘I still don’t see them.’”

- Theme 2: text-based communication challenges
  - 2a: feels impersonal
    - “Maybe just that it’s less personalized...you don’t get to see a face on the telephone but somebody just talking to a voice... especially if you’re anxious about your care, you want reassurance that you did the right thing.”

  - 2b: difficulties in relaying the uniqueness of a patient
    - “You lose the sort of nuances of the, of the patient and...to think about if you knew a little bit more about...the background of the patient or the story.”

  - 2c: miscommunications may occur
    - “It’s nice that things are documented, but sometimes things are missed in the documentation. And so, you’re making clinical decisions or clinical consultation suggestions based off of someone’s assessment that may or may not be correct.”

  - 2d: does not provide opportunities for back-and-forth dialogue
    - “So I think that that sort of in-time back and forth and counseling can be really valuable...as opposed to...the written word”

  - 2e: limits opportunities for network building
    - “You miss out on some of the networking, like, some of the personal and interpersonal dialogue that sometimes helps relationships grow, or trust grow.”

- Theme 3: not receiving an answer in real time
PCPs noted that unlike telephonic consultation, electronic consultation gave them the flexibility to submit the consultation question and review the response at times that were convenient for them, which was especially helpful in ensuring that the consultation did not detract from the existing patient care responsibilities. PCPs also noted the benefit of receiving timely specialist recommendations in writing, which is not often possible with telephonic consultations. Similarly, the participants found the documentation of both their consultation and the specialist’s response in the EHR to allow them to refer back to it in the future, should a similar question arise for another patient, to be particularly helpful. In addition, a few providers noted that the electronic consultation had the potential to allow both themselves and the specialist receiving the consult to be reimbursed for their time, which is not possible with telephonic consultations. Finally, some PCPs noted feeling that the act of submitting an electronic consultation would help them to better communicate their clinical questions to the specialist:

I think that sometimes being able to put it down and refine it, like, ‘No, no, this is my question, and this is my patient,’ before you send it off has significant value, because then it helps you sort of integrate and synthesize before sending it off.

Regarding the drawbacks of electronic consultations, five themes emerged: (1) EHR-related difficulties, including EHR incompatibility and unfamiliarity; (2) text-based communication challenges; (3) not receiving an answer in real time; (4) forced conversations with patients about billing; and (5) difficulties for providers who lack baseline knowledge regarding gender-affirming care (Textbox 2). Multiple PCPs cited concerns about being unfamiliar with the EHR used by specialists in their area and that making an electronic consultation system available only to those who use a specific EHR would make it inaccessible for many PCPs. In addition, specific concerns arose about the text-based electronic consultation communication, which some felt could feel impersonal, make it challenging to relay the specific nuances of a case, or lead to miscommunications between providers. The participants also discussed concerns that electronic consultation may not provide opportunities to engage in back-and-forth dialogue with a specialist, as opposed to telephonic consultation, and regarding limited opportunities for networking with other providers as would be possible with tele-education. Another concern reported by PCPs regarding electronic consultation was not having the ability to receive an answer to their clinical question in real time as they would be able to do with telephonic consultation. Finally, some providers expressed discomfort with the idea of having to inform patients and families that they would be billed for the electronic consultation. This was illustrated well by a provider who had previously used an electronic consultation platform for psychiatry:

I hate talking about money, right? I just want to take care of patients. So I expected [talking about billing for the electronic consultation with a psychiatrist] to be a very uncomfortable conversation where I say... “You know I can reach out to our pediatric psychiatrist, but this is a special consult and it will be billed to your insurance.” And I just felt kind of gross and icky, it’s almost like the family...feels like they have to say yes.

Finally, a few PCPs noted that they felt that electronic consultation would be most useful if they had strong foundational knowledge regarding gender-affirming care, which they could receive through other continuing medical education, such as the tele-education platform.

Telephonic Consultation

With respect to the advantages of telephonic consultation, three themes emerged (Textbox 3): (1) having a dialogue with a specialist, (2) receiving compensation for PCPs’ time, and (3) helping with acuity or complex cases.
Textbox 3. Themes primary care providers identified as the benefits and drawbacks of telephonic consultation.

**Benefits**

- **Theme 1: having a dialogue with a specialist**
  - 1a: can ask additional clarifying questions
    - “Talking with someone over the phone, sometimes it’s beneficial because they’ll ask follow-up questions that you didn’t ask that can be a learning tool, but then also identify, maybe, some blind spots that maybe should be identified before people provide a specific answer, which I think is somewhat of a safety net for catching some of the clinical biases that we might have in medical decision making.”
  - 1b: can relay nuances of the patient’s situation
    - “I think the obvious benefits again are timeliness and being able to sort of convey the nuances of the story, or the patient. I think that having a conversation is better than a template when you’re talking about patients.”
  - 1c: already comfortable using this modality
    - “I think one of the benefits is it’s a model we’re familiar with and we already do it, and so it seems pretty easy to be able to, you know, call the Children’s provider to provider line, and now I can ask for a gender specialist, instead of just an endocrinologist”
  - 1d: receive a response in real time
    - “When we’re in conversation with families, we can let them know, ‘Hey, I don’t have an answer to your question right now, but I know who to call, and I know that they’re going to get back to me by five o’clock and then, therefore, I will get back to you today or tomorrow morning.’”

- **Theme 2: receiving compensation for primary care providers’ time**
  - “As our coders and billers have told us...if we do the consult the same day that’s part of our coding to have for the visit, and so it could be, you know the charge can be captured in that sense as well.”

- **Theme 3: helping with high acuity or complex cases**
  - “I think this one would be better for those more, like you said, life and death situations. Or more severe. Like, I don’t want to, maybe they’re not like actively suicidal and I don’t need to send them to the emergency department, but, like, I’m very worried about them and I don’t want to wait 24-48 hours to hear back.”

**Drawbacks**

- **Theme 1: challenges with using the phone for communication**
  - 1a: the timing of callbacks is unpredictable and may be inconvenient
    - “It is hard when you call and [the specialist] is going to call back at the end of the day, with time zone differences. I mean, I’m not always still at work, and then if I have my cell phone, it feels like I’m on call because I want to be respectful of [the specialists’] professional time. But sometimes it’s really more disruptive, because I’m not in front of a chart or things when I get the callback. Like, out walking the dog or with kids, just other responsibilities.”
    - “If I’m seeing 24 patients in a day [and] you put in the stress of...Children’s is going to call me back and the MA is going to pull me from the room and the family’s pissed because I was already 25 minutes late for them. Yeah now I’m getting pulled out and...am stressed about the backup that’s happening.”
  - 1b: no visual record of specialists’ recommendations
    - “Sometimes over the phone, you are scrambling to write some sort of notes or maybe write down the number...”
  - 1c: difficulty in relaying necessary data
    - “I think another drawback, if there’s a way to be able to send this stuff, like, electronically, you know, like labs and things like that. If the person wanted to see them, you can just imagine, like, rattling them off to the poor person trying to help you and I’m like, ‘hold on,’ you know? So, having that visual component...would be missed in the telephone one.”
  - 1d: phone calls can be intimidating and awkward
    - “It’s a little intimidating to call somebody even if they’re...super nice. It can be a little, like, they’re going to think...I’m dumb, and, you know, kind of a...med student kind of feel, you know?”

- **Theme 2: limited expertise of the responding providers**
  - “It would be nice to have a direct line to the gender clinic, so I know that the provider that I’m, that I’m paging is specifically that.”
More specifically, PCPs felt that telephonic consultation allowed them to ask clarifying questions and convey subtleties of the patient’s case, which may be difficult to communicate via text-based methods such as electronic consultation. In addition, PCPs reported that the timeliness of the consultation (within 24 hours) was very reassuring and allowed them to provide a timely response to a patient or their family instead of waiting days for an electronic consultation or even weeks for tele-education. Finally, PCPs noted increased comfort in using telephonic consultation platforms, given that it is a model that many had previously used.

With respect to the drawbacks of telephonic consultation, 3 themes emerged (Textbox 3): the (1) challenges associated with using the phone for communication, (2) limited expertise of the responding providers, and (3) lack of utility for nonemergent issues. PCPs cited facing multiple logistical challenges, leading to frustration with phone-based communication, including receiving return phone calls at inconvenient times:

[We have] no idea what window the person is going to call us back in and so if we’re in the middle of a very difficult discussion with a family in a room having someone come and knock on the door and say ‘you have a phone call’ is very disruptive.

They also reported difficulties with not having a specialist’s recommendations documented in writing and challenges relaying necessary data (eg, laboratories) accurately over the phone.

Some PCPs noted that they found it intimidating or awkward to make calls to specialists, particularly to individuals with whom they did not have a relationship. Similarly, they found it cumbersome not to have a direct line of communication with a specialist who has specific experience in gender care:

I’ve gotten...an adolescent medicine provider who’s more specialized in something else, like eating disorders, or menorrhagia...and then it’s a lot of back and forth, you know? Or it’s like, ‘Oh, let me go talk to my attending about that,’ and then they...go and then they come back, or call me again later and I’m in the room.

Some PCPs also reported feeling that most consultation questions that arose regarding their TGD patients did not feel urgent or time sensitive enough to warrant a same-day response:

I usually text or email or do something like that...Just because I didn’t need the answer right away. And some of it was to...solidify knowledge...and some of it was patient care; then I would call back the family in the next day or two or the patient in the next day or two.

Most Effective Platform

PCP perspectives regarding which platform would be most effective in supporting them in delivering gender-affirming care in the pediatric primary care setting were relatively mixed, with nearly equivalent numbers of providers preferring tele-education, electronic consultation, and telephonic consultation. In addition, multiple PCPs indicated a preference for using a hybrid of the 3 models, citing that a single platform alone may not be sufficient to support them in delivering gender-affirming care (Table 1).

- “Sometimes I got people and they were like ‘Oh, I don’t know that, let me get this provider to call you back,’ and then so it ended up resulting sometimes in a couple of phone calls.”

- Theme 3: lack of utility for nonemergent issues
  - “I have to say...I try to be cautious about paging...just because I feel like [specialists] are so busy, right? And I think most of the gender stuff is not urgent or emergent...not even end of day kind of questions and so, although I like...having the ability to [use telephonic consultation]....I don’t know that I would use it.”
Table 1. Primary care providers’ preferred modality to provide support in delivering gender-affirming care in the general pediatric setting (N=15).

<table>
<thead>
<tr>
<th>Preferred platform</th>
<th>Participants, n (%)</th>
<th>Characteristics of a successful platform</th>
<th>Representative quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Telephonic consultation</td>
<td>3 (20)</td>
<td>Nonjudgmental approach and timeliness</td>
<td>• “The most useful...would probably be the telephone consultation. I think I would use it the most, and I think it would...have the most impact.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tele-education</td>
<td>4 (27)</td>
<td>Sense of community, practicality, ability to engage users, and comprehensiveness</td>
<td>• “Probably the ECHO program because it’s a sustainable teaching method whereas the e-consult and the telephone, there may be some teaching involved but it’s essentially just giving you the answer. Which then tells you how to help this one person, but it may help you with a few others, but it’s really, just, very, it’s very individualized for the person in front of you. The ECHO program, not to use analogies too much, but it’s sort of the, you know, you give the person a fish, you feed him for a day, but if you teach them how to fish they’ll be able to feed themselves for the rest of their lives. So, I think, in the end, the one which is going to be the most beneficial is going to be the ECHO program.”</td>
</tr>
<tr>
<td>Electronic consultation</td>
<td>4 (27)</td>
<td>Reliability, technological accessibility, timeliness, and scalability</td>
<td>• “I think that at the end of the day...I’d probably go with electronic consultation because it allows me as a medical provider the most flexibility. I can send that message at 7pm or 4am when I’m writing notes, as opposed to...being limited to the scope of...on your lunch hour or...within the business day.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• “I think electronic consultation sort of allows for the greatest synthesis of assessment and I think that can be valuable to everyone...I need to write, send my question,... and then have that come back as a response, is probably the most valuable thing because sometimes you just need to think about it before you, sort of, ask that question.”</td>
</tr>
<tr>
<td>Hybrid</td>
<td>4 (27)</td>
<td>Flexibility, scalability, comprehensiveness, adaptability, timeliness, and integration of different modalities</td>
<td>• “I think none of these modalities would likely be...enough on their own, right? Like, I think...in an ideal state, you would have multiple ways of communicating depending on the intensity of what’s going on. If I need to really talk to someone right now about something really intense happening with a patient right now, they’re having a severe reaction to some medicine that someone else has provided that I don’t really know about, I need to talk to them right now, right? And within 48 hours is not okay. But other things where...I have...more general questions or decisions that need to be made over some weeks, then doing them electronically is great.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• “I really think it may have to be all three...the TeleECHO, I definitely don’t think will be enough regarding specific patients. The electronic consultation, you could do it that way, but I think you get more information if you had some ECHO too. And phone consultation just again, unless they’re really looking at their electronic stuff on a very...regular basis, sometimes you just gotta reach out and say...’Do you understand what I’m saying?’...or it’s too much! Like the kid with the psych stuff, there was a lot of stuff and I just kind of wanted to say... These have changed, this is what’s going on, this is why this is like this now, this is what I’m thinking,’ you know? And I didn’t want to write a two page letter.”</td>
</tr>
</tbody>
</table>
Characteristics of a Successful Platform

The 10 characteristics PCPs felt were most important for the success of platforms for supporting PCPs in providing gender-affirming care to TGD youths are shown in Table 2.

### Table 2. Characteristics pediatric primary care providers (PCPs) felt were most important for the success of a platform for facilitating gender-affirming care provision in the primary care setting.

<table>
<thead>
<tr>
<th>Platform characteristics</th>
<th>Representative quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reliability</td>
<td>“If you guys say it’s 24-48 hours and people are not responding to me for a week, I’m going to stop. I’m going to just start using the telephone instead.”</td>
</tr>
<tr>
<td>Timeliness</td>
<td>“Getting things done quickly and being able to get back to either the family or the youth quickly, to figure out the next step. I find that if I let things sit for too long, things fall through the cracks.”</td>
</tr>
<tr>
<td>Ability to engage users</td>
<td>“Making sure to involve everybody, and you know, keep them accountable. Keeping people engaged would be huge.”</td>
</tr>
<tr>
<td>Scalability</td>
<td>“I think also knowing that if there were a case that anybody from the gender clinic thought the patient could use a higher level of care, if it were an option to have the patient do a consultation with somebody in the clinic would be cool.”</td>
</tr>
<tr>
<td>Practicality</td>
<td>“Sometimes we have [specialists] come in and...talk at a level that is, like, what they would talk to their colleagues. And I already don’t know what you’re talking about. I’m confused. And so, getting some baseline understanding that we are primary care providers...Understand that there’s going to be, terminology that you use...that is already going to be above us. And, all providers, no matter how much we want to pretend that we don’t, we’re all a little bit proud. And so, it’s hard for us to be, like, ‘I don’t know what that means.’ And so, it would be like having that understanding [that] there’s a lot that we don’t know. And so, if they were to just start, rattling off about, ‘Here’s dosages,’ I’d be like, ‘Whoa, hang on a second. Let’s go back. Which ones are for boys and which ones are for girls?’ I think that bringing it down to the primary care level to start is important.”</td>
</tr>
<tr>
<td>Adaptability</td>
<td>“I think, you know, changing with the times. Because, all of this information, I feel like it’s constantly changing. I’m always reading about, new terminology, new ways that people like to be referring to, new ways that that you’re supposed to ask questions. So, changing with that and letting us know that those are changes. Because sometimes we don’t even know. We’re like, ‘Oh, is that the way it’s always been? Cool.’ And [also] teaching us. ‘Hey, this is going to be one of those things where, every single time we talk it’s going to be slightly different.’”</td>
</tr>
<tr>
<td>Comprehensiveness</td>
<td>“Any system has to be comprehensive if there’s going to be an ask for me and other PCPs to do more than what we’re doing right now. And to move on to [prescribing gender affirming medications] and [referring patients for gender-affirming surgeries] it is really going to take a lot of support, because we just don’t have experience [to know] when to pull the trigger.”</td>
</tr>
<tr>
<td>Integration of different modalities</td>
<td>“Having some sense of connectedness between them. So, if you had the capacity to say here’s my submission of my consultation, and if more information is needed, I’m happy to talk on the phone about this.”</td>
</tr>
<tr>
<td>Accessibility</td>
<td>“Different EMRs and making sure there’s some way to adapt to people who don’t have Epic.”</td>
</tr>
</tbody>
</table>

Specifically, PCPs desired platforms that provided reliable and timely consultative support as well as those that were engaging and nonjudgmental. In addition, the participants indicated that successful platforms were those that provided practical information, could scale up to having the specialist conduct a formal consultation with the patient if necessary, and could further adapt as terminology and pediatric gender-affirming care delivery evolve. Finally, PCPs desired platforms that were comprehensive, integrated, and accessible to all providers, and not just those using a specific EHR.

When asked what would make them more likely to continue using a particular platform, PCPs cited patient satisfaction, community building, and incentives for participation. First and foremost, PCPs frequently indicated that being able to improve the care they provide to TGD youths would motivate their use of these platforms:

*I think the biggest thing for me is...patient satisfaction. Right? If it felt like the right thing to do for the patient and the patient was happy, grateful, thankful, whatever you want to say, right? Like, if it’s...* Oh, I
can...stop your periods today.” That makes a huge difference for a 14-year-old, who...was born female and is a male, and is distressed every single month. So to be able to...help that kid would make me do it again.

Providers also felt that having opportunities to establish a sense of community with other pediatric providers interested in gender care would motivate them to continue using a platform:

Certainly, hearing from other providers having similar experiences...would make me want to go back and have another...tele-education meeting. Just to know that I’m able to glean information for my patient care from those meetings.

Finally, PCPs mentioned that being able to receive incentives, such as continuing medical education or the maintenance of certification credits, would motivate them to use a platform. This was especially true for tele-education platforms, as many participants acknowledged that such incentives could offset the significant time commitment required for participation.

**Discussion**

**Principal Findings**

The results of this qualitative study suggest that pediatric PCPs desire opportunities to both obtain foundational knowledge and receive timely consultative support from gender specialists regarding patient-specific concerns. The variation in these results stems from the wide variation in PCPs’ training and experience in providing gender-affirming care for TGD youths, and the current options for training and consultation in this area are quite limited [13-15]. To meet the increasing demand for gender-affirming care services for TGD youths, we must develop a diverse suite of telehealth-based training and consultation services to meet the needs of PCPs with different levels of experience and training in this area. This specialist-to-PCP support is critical for facilitating greater PCP independence in gender-affirming care provision as well as for expanding the access of TGD youths to pediatric gender-affirming care services.

Increasing requests for specialist-to-PCP telephonic consultation during the COVID-19 pandemic [31] have led many large pediatric health systems to reconsider whether these services are (1) providing the best quality care to patients and (2) sustainable for pediatric specialists [32,33]. Although informal telephonic or “curbside” consultations remain the most common form of pediatric specialist-to-PCP telehealth support, our findings suggest that it may not be the ideal modality to support PCPs in providing pediatric gender-affirming care. Although our findings indicate that PCPs perceived telephonic consultation as having some important benefits, such as the timeliness of response and wider accessibility, they also noted many drawbacks. These drawbacks, which include limited compensation for consultation services, raise concerns about the sustainability of telephonic consultation systems and indicate a need to develop new modalities to provide specialist-to-PCP support.

Consequently, several providers in our study expressed a desire for an electronic consultation platform to support the provision of gender-affirming care in the pediatric primary care setting. This modality may be particularly useful in overcoming some of the barriers that exist with telephonic consultation systems, including the lack of written documentation, inconvenience of receiving unscheduled phone calls, difficulty in exchanging laboratory data, and lack of PCP and specialist compensation. In particular, electronic consultation may be a helpful modality to increase the capacity of PCPs to submit nonurgent questions to support their TGD patients and to ensure that questions are routed directly to providers with expertise in gender-affirming care. Electronic consultation may also increase the capacity of PCPs and specialists to exchange comprehensive and patient-specific information, review objective data, and document recommendations in writing to facilitate the provision of ongoing care.

Despite these benefits, there are some challenges to developing specialist-to-PCP electronic consultation platforms. Several PCPs indicated that they desired an electronic consultation platform within the EHR used in their practice because of both its convenience and their familiarity with its functionality. However, this remains logistically challenging given the heterogeneity of EHRs used by pediatric PCPs across the United States and the reluctance of EHR vendors to adopt sustainable medical applications, reusable technologies application programming in accordance with defined standards for Fast Healthcare Interoperability Resources [2,34]. This is likely why many of the existing specialist-to-PCP electronic consultation platforms, such as those used by the Veteran Affairs health system [16,17,21], are available to only PCPs who work within the same health system as that of the specialists providing the consultation. Ensuring that pediatric gender-affirming care provision is accessible and equitable will require the use of modalities that are widely accessible to providers in diverse clinical practice settings [35].

Tele-education may also be a particularly useful modality for PCPs whose practices are located farther from a pediatric gender specialist or who are seeing an increasing number of TGD patients [18]. Regarding geography, providers located farther from pediatric multidisciplinary gender clinics may be more inclined to dedicate time to formal education sessions, as they serve patients who face additional access- and travel-related barriers to receiving specialty care. Given that this platform would provide them with an opportunity to receive more comprehensive foundational knowledge, providers in remote areas may be more willing to invest time upfront, despite the clinical sacrifices, knowing that it would facilitate care for their patients. Finally, patient volume, specifically, the number of TGD youths seen in their practice, may impact their interest in a tele-education platform. On the one hand, the increasing number of TGD patients may encourage PCPs to gain more formal experience working with this population; alternatively, a PCP who already sees many TGD youths in their practice could be more inclined to use telephonic or electronic consultation, as they are more likely to have developed foundational knowledge and skills through practice.
Limitations
This study should be interpreted within the context of the following limitations. Although diversity existed with respect to participants’ primary practice locations and years in practice and the number of patients they referred to a gender clinic, the participants were relatively homogeneous with respect to gender identity and race and ethnicity. In addition, both our response rate and decision to recruit PCPs who had previously sought support may limit the generalizability of our findings. Although the providers in our study may be more likely to use a telehealth platform for support in providing gender-affirming care than those who have not sought out this support, we are confident that these data reflect the perspectives of PCPs who are the most motivated to use a telehealth platform for support in providing care for TGD youths in the primary care setting. Furthermore, although our interview guide was intentionally designed to obtain PCPs’ perspectives about both the advantages and disadvantages of each modality, it is possible that social desirability bias affected our findings. Finally, given that specialist-to-PCP telephonic consultation systems are currently in use in many pediatric hospital systems, it is likely that the PCPs in our study had more experience using this modality than electronic consultation or tele-education, which could, in turn, have affected their responses.

Conclusions
In summary, our findings suggest that beyond the current telephonic consultation model, electronic consultation and tele-education may provide important alternative training and consultation platforms to support pediatric PCPs in providing gender-affirming care to TGD youths. Improving specialist-to-PCP support in these ways is critical for facilitating greater PCP independence in gender-affirming care provision and promoting widespread access to pediatric gender-affirming care services for TGD youths.

Acknowledgments
This project was supported by the Seattle Children’s Research Institute Career Development and Agency for Healthcare Research and Quality K12HS026393-03 (principal investigator: GMS).

Authors’ Contributions
GMS, NFK, PGA, WP, TS, DAC, and LPR provided substantial contributions to the conception or design of the work. KMB, GMS, and NFK provided substantial contributions to the acquisition, analysis, or interpretation of data for the work. KMB, GMS, NFK, PGA, and TS were involved in drafting the manuscript. WP, DAC, LPR, GMS, and NFK were involved in critically revising the manuscript for important intellectual content. All the authors provided final approval of the manuscript version to be published and agreed to be accountable for all aspects of the work and involved in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Definitions used to describe each telehealth platform to the interview participants.

References


Abbreviations

- EHR: electronic health record
- PCP: primary care provider
- SCGC: Seattle Children’s Gender Clinic
- TGD: transgender and gender-diverse

©Gina M Sequeira, Nicole F Kahn, Kevin M Bocek, Taraneh Shafii, Peter G Asante, Dimitri A Christakis, Wanda Pratt, Laura P Richardson. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 31.01.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Designing a Future eHealth Service for Posthospitalization Self-management Support in Long-term Illness: Qualitative Interview Study

Hege Wathne1*, MSc; Ingvild Margreta Morken2,3*, PhD; Marianne Storm4*, PhD, Prof Dr; Anne Marie Lunde Husebø1,3*, PhD, Prof Dr

1Department of Public Health, University of Stavanger, Faculty of Health Sciences, Stavanger, Norway
2Department of Quality and Health Technologies, University of Stavanger, Stavanger, Norway
3Research Group for Health and Nursing Sciences, Stavanger University Hospital, Stavanger, Norway
4Faculty of Health Sciences and Social Care, Molde University College, Molde, Norway
*all authors contributed equally

Corresponding Author:
Hege Wathne, MSc
Department of Public Health
University of Stavanger
Faculty of Health Sciences
Kjell Arholms gate 41
Stavanger, 4036
Norway
Phone: 47 98019955
Email: hege.b.wathne@uis.no

Abstract

Background: For patients with noncommunicable diseases (NCDs; eg, heart failure [HF] and colorectal cancer [CRC]), eHealth interventions could meet their posthospital discharge needs and strengthen their ability to self-manage. However, inconclusive evidence exists regarding how to design eHealth services to meet the complex needs of patients. To foster patient acceptability and ensure the successful development and implementation of eHealth solutions, it is beneficial to include different stakeholders (ie, patients and health care professionals) in the design and development phase of such services. The involvement of different stakeholders could contribute to ensuring feasible, acceptable, and usable solutions and that eHealth services are developed in response to users’ supportive care needs when transitioning to home after hospitalization. This study is the first step of a larger complex intervention study aimed at meeting the postdischarge needs of 2 NCD populations.

Objective: This study aimed to explore the perspectives of patients with HF and CRC and health care professionals on patient self-management needs following hospital discharge and investigate how a future nurse-assisted eHealth service could be best designed to foster patient acceptability, support self-management, and smooth the transition from hospital to home.

Methods: A qualitative, explorative, and descriptive approach was used. We conducted 38 semistructured interviews with 10 patients with HF, 9 patients surgically treated for CRC with curative intent, 6 registered nurses recruited as nurse navigators of a planned eHealth service, and 13 general practitioners experienced in HF and CRC treatment and follow-up care. Patients were recruited conveniently from HF and CRC outpatient clinics, and the nurses were recruited from the cardiology and gastro-surgical departments at a university hospital in the southwest of Norway. The general practitioners were recruited from primary care in surrounding municipalities. Semistructured interview guides were used for data collection, and the data were analyzed using thematic analysis.

Results: In total, 3 main themes were derived from the data analysis: expecting information, reassurance, and guidance when using eHealth for HF and CRC self-management; expecting eHealth to be comprehensible, supportive, and knowledge promoting; and recognizing both the advantages and disadvantages of eHealth for HF and CRC self-management. The data generated from this interview study depicted the diverse needs for self-management support of patients with CRC and HF after hospital discharge. In addition, valuable suggestions were identified regarding the design and content of the eHealth service. However, participants described both possible advantages and disadvantages of a remote eHealth service.
Conclusions: This study is the first step in the development of an eHealth service for posthospitalization self-management support for long-term illnesses. It concerns patients’ supportive care needs and user requirements of an eHealth service. The findings of this study may add value to the planning and development of eHealth interventions for patients with NCDs.

(JMIR Hum Factors 2023;10:e39391) doi:10.2196/39391

KEYWORDS

colorectal cancer; eHealth service; heart failure; noncommunicable diseases; self-management; qualitative research; mobile phone

Introduction

Background

Noncommunicable diseases (NCDs) are defined as diseases or conditions that tend to be of long duration and slow progression [1]. NCDs are estimated to be responsible for >70% of all deaths (41 million people) per annum worldwide, and the most common NCDs that account for the most deaths are cardiovascular disease and cancer [2]. A growing number of patients with NCDs such as heart failure (HF) and colorectal cancer (CRC) are prone to comorbidities, a high rate of readmissions, and complex health care needs [3]. Similar to all long-term chronic conditions, patients may require day-to-day self-management [2], which may result in increased treatment burden (ie, patient work) [3,4]. They also experience an ongoing need for a trustworthy contact with the health care system that can deliver qualitatively sound health-related information [4,5].

HF is a progressive and complex clinical syndrome with a tremendous symptom burden, including dyspnea, fatigue, edema, and sleeping difficulties [6,7]. It is associated with periods of acute deterioration and an increased risk of hospitalization [8,9]. CRC is one of the most prevalent cancers in the world [10]. Owing to improvements in health care systems, the number of survivors of this cancer has increased [11]. This causes patients with CRC to live with the illness for a longer period, similar to patients with other chronic diseases [12].

The Importance of Posthospitalization Self-management

The period following hospital discharge is deemed particularly vulnerable for many patients as they transition from care in a safe hospital setting to individual self-care at home [13]. Moreover, many struggle to perform recommended self-care and navigate the health care system, particularly when posthospitalization care is poorly executed because of inadequate coordination of resources or follow-up from home health care interventions or general practitioners (GPs) [4,9].

Self-management may be defined as the strategies that individuals undertake to promote health, manage an illness, and manage life with an illness [14]. Self-management is increasingly recognized as a fundamental component of NCD care as adequate self-management skills may help patients with NCDs control their chronic conditions [15]. However, self-management demands a substantial effort from the patient, requiring routine work and timely adjustment of therapy to avoid exacerbation events and facilitate detection and avoidance of recurrence and prevention of disease progression. In addition, patients must solve practical problems, manage physical and psychosocial consequences and lifestyle changes, and know when and how to seek appropriate medical advice [16-18]. In HF management, self-care is a cornerstone as it improves treatment effectiveness and reduces hospital admissions. However, many patients with HF have a limited understanding of the basic elements of the nature of HF; they often misinterpret HF symptoms and feel inadequately informed [19]. Consequently, patients with HF are often unprepared to take charge of their self-management tasks after hospital discharge [19]. Therefore, self-management interventions that promote and support self-care after hospital discharge are becoming increasingly important for this group of patients [19]. In patients with CRC, a decrease in postoperative length of stay has been observed [20]. However, many patients are likely to experience changes during the initial postoperative phase, including changes in bowel habits, pain, fatigue, mobilization, dietary challenges, and physical and psychological distress (ie, anxiety and depression) [21]. Many patients with cancer also experience ongoing difficulties in assessing support and services at home [22]. Hence, the transition from active treatment in the hospital to self-care at home is a period when patients with CRC most feel insecure and require intervention [12].

eHealth and Current Self-management Programs

Today, health care systems worldwide are faced with the challenge of managing care for long-term chronic illnesses [23]. An extraordinary and promising resource that promotes chronic disease management, including patient self-management, is eHealth [24,25]. eHealth is defined as the delivery of health care using modern electronic information and communication technologies when health care providers and patients are not directly in contact and their interaction is mediated by electronic means [25]. The purpose of eHealth is to change patients’ behavior and improve their health status [26]. eHealth may also enhance treatment durability as patients can receive support and reinforcement of skills after hospitalization during the transition phase from hospital to home [27]. Research suggests that patients with chronic illnesses supported by innovative eHealth solutions within a care pathway feel more motivated to engage in self-management behavior [28,29]. Furthermore, a study investigating video consultation as an alternative to face-to-face consultation among patients with CRC and their treating surgeons showed that video consultation is equivalent to face-to-face follow-up consultations in terms of patient satisfaction and perceived quality of care [30]. This may suggest that the quality of patient-provider interaction can be maintained using digital solutions [31]. However, self-management support interventions need to be tailored to the individual and their specific condition and context [32]. Moreover, the growing number of patients with NCDs requires a more dynamic and flexible follow-up approach, and eHealth support may be a...
beneficial strategy to meet the posthospital discharge needs of patients with NCDs [3,24].

**Fostering Patients’ eHealth Acceptance**

The potential benefits of eHealth have been widely described. However, the use of eHealth remains low, and evidence on how to design eHealth services to meet the complex needs of patients is inconclusive [33]. An explanation for the lack of results is that, during the development process, insufficient attention is paid to the needs, wishes, and context of the prospective end users [32]. The credibility, value, and success of eHealth lie in its ability to demonstrate positive outcome effects, where end users’ engagement in the design and development of eHealth services is important to overcome adaptability barriers [34]. Today, several studies have highlighted the lack of user involvement in the development of such interventions [35-38]. To foster the successful use of eHealth interventions, it is important to develop eHealth interventions in response to users’ needs rather than as a technological innovation [37], and for self-management support to be effective, it must be provided by suitable health care professionals (HCPs) [39]. In particular, nurses are important to support self-management as enabling patients to understand and cope with their disease, its treatment, and its consequences is a core competence of nursing [39]. Nurses can, through remote digital care, guide and support patients in self-care by providing them with analytic skills to interpret bodily signals and by activating them to take the appropriate measures to prevent exacerbation events [40]. Therefore, nurses may play a pivotal role in fostering patient acceptance and support and guiding patients toward sustainable and effective self-management [40].

**Current Knowledge Gap and the Need for This Study**

Research on eHealth-based support interventions for people with NCDs recommends that the interventions be theory-based and hold an element of communication in addition to web-based material [41]. eHealth programs are found to be most efficient when led by multidisciplinary teams where HCPs can encourage the patients to adhere to the program and when the eHealth program is designed based on the outcomes to be achieved [42]. Critical gaps remain in the design and evaluation of self-management interventions, with a lack of patient and clinician involvement [43]. Rochat et al [44] emphasized the importance of iterative involvement of end users in the design and evaluation process of a coaching solution to support the postdischarge needs of patients with HF. Furthermore, the results of Fairbrother et al [45] showed that telemonitoring enhanced patients’ knowledge and understanding of their condition but that further work is required by patients and professionals to develop a shared understanding of self-management and the role and function of telemonitoring as an enabling intervention within this context.

Although appearing different in terms of diagnosis, treatment, and prospects, patients with HF and patients surgically treated for CRC both represent conditions in need of long-term follow-up care, necessitating extensive self-management capacity and skills in the transition to home after hospital discharge [46,47]. Moreover, the 2 patient groups have the most vulnerable types of NCDs and may serve as proxies for the broader NCD field. Self-management interventions across different chronic conditions can contribute to improved health outcomes [48]. A recent study found that survivors of CRC were positive toward postdischarge monitoring and follow-up. The participants especially requested features for information, questions and answers regarding nutrition and weight, and provision of social support [30]. Research on digital self-management interventions for patients with HF has shown varied results [43,49]. When used for posthospitalization follow-up, eHealth interventions can positively affect quality of life, whereas their impact is less evident for self-care and readmissions [49]. However, research on how patients with NCDs can best be supported in self-management during transitions is sparse, including which eHealth-based support interventions are best suited for follow-up care [13]. To many patients with HF or CRC, the transition to self-management after hospital discharge represents a void of professional health care that may leave them unprepared for self-managing these tasks at home [50,51]. Thus, bridging the gap in health care between hospital discharge and home by developing more seamless eHealth services from inpatient to outpatient care supported by hospital assistance seems necessary if patients with NCDs are to achieve adequate self-care and feel safe [6,16].

**Aims of This Study**

In this study, which is the first step of a larger complex intervention aimed at developing and testing a generic eHealth service for patients with NCDs, the aims were twofold: (1) to explore the supportive care needs of patients with HF and patients surgically treated for CRC in transition to home and (2) to identify different stakeholders’ (ie, patients, registered nurses [RNs], and GPs) views on important content and functions of a future eHealth service designed to meet patients’ supportive care needs in the transition from hospital discharge to home. The research questions were as follows: (1) What are the essential needs regarding self-management support among patients with HF and CRC transitioning from hospital to home that can be met by a future eHealth service? (2) How can a future eHealth service be best designed, and what are perceived to be essential content and functions to foster patient acceptability from the perspective of patients and HCPs?

**Methods**

**Study Setting and Design**

This study is part of a larger research project, eHealth@Hospital-2-Home, and includes three phases: (1) developing a nurse-assisted eHealth service, (2) assessing feasibility and piloting the service, and (3) carrying out a randomized controlled trial [3]. This study pertains to the main project’s first phase and will inform the modeling and adaption (ie, content and functions) of a future hospital-based, nurse-assisted eHealth service for patients living with HF or CRC. In this study, an exploratory and descriptive qualitative design was applied. Data were collected using semistructured interviews with patients with HF, patients surgically treated for CRC, RNs, and GPs to explore their perspectives on patients’ supportive needs following hospital discharge and how a future
eHealth service can be best designed to foster patient acceptability.

**Study Population**

The study’s patient populations comprised patients with HF and patients who had received surgical treatment for CRC with curative intent. The selection criteria for both patient groups were age between 18 and 80 years, attendance to an outpatient clinic at hospital A, ability to understand and speak Norwegian, capability to take part in the interview, and no acute medical crisis. The patients were recruited during a scheduled follow-up appointment at either the HF outpatient clinic or the gastro-surgical outpatient clinic. The study sample also comprised nurses and GPs. The nurses, engaged as nurse navigators (NNs) in the project, were RNs from 2 hospitals in the southern part of Norway (hospitals A and B) and experienced with HF or CRC treatment. In total, 4 of the RNs worked at hospital A: 2 in a medical intensive care unit in the cardiology department and 2 in a gastro-surgical ward. The final 2 RNs worked in an HF unit at hospital B. The GPs worked as part of primary care services in municipalities corresponding to hospital A, and they all had ≥2 years of experience as GPs. The GPs were invited to participate in the study because of their experience with various patient groups, including HF and CRC, after hospital discharge.

A total of 39 persons were approached, and 38 (97%) consented to participate in the study. Of these 38 participants, 10 (26%) were patients with HF, 9 (24%) were patients surgically treated for CRC, 6 (16%) were NNs, and 13 (34%) were GPs. The age of the patients with HF ranged from 49 to 78 years, and that of the patients with CRC ranged from 58 to 76 years. The RNs were all women and ranged in age from 26 to 37 years. Their work experience as RNs ranged from 3 to 11 years, and 67% (4/6) of the nurses were nurse specialists (ie, intensive care and stoma nurses). The GPs were aged 35 to 66 years, and their work experience as medical doctors was between 7 and 27 years. All but 23% (3/13) of the GPs were specialized in general medicine, as well as 15% (2/13) who were also specialized in community medicine. Please see Table 1 for an overview of participant demographics.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Patients with CRC(^a) (n=9)</th>
<th>Patients with HF(^b) (n=10)</th>
<th>GPs(^c) (n=13)</th>
<th>NNs(^d) (n=6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), range</td>
<td>58-74</td>
<td>49-73</td>
<td>35-66</td>
<td>26-34</td>
</tr>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>3 (33)</td>
<td>7 (70)</td>
<td>9 (69)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Female</td>
<td>6 (67)</td>
<td>3 (30)</td>
<td>4 (31)</td>
<td>6 (100)</td>
</tr>
<tr>
<td><strong>Educational status of patients, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary school</td>
<td>3 (33)</td>
<td>0 (0)</td>
<td>N/A(^e)</td>
<td>N/A</td>
</tr>
<tr>
<td>High school</td>
<td>5 (56)</td>
<td>6 (60)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>College or university</td>
<td>1 (11)</td>
<td>4 (40)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Work experience for GPs and NNs (years), n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-3</td>
<td>N/A</td>
<td>N/A</td>
<td>1 (8)</td>
<td>1 (17)</td>
</tr>
<tr>
<td>4-7</td>
<td>N/A</td>
<td>N/A</td>
<td>2 (15)</td>
<td>4 (67)</td>
</tr>
<tr>
<td>&gt;10</td>
<td>N/A</td>
<td>N/A</td>
<td>10 (77)</td>
<td>1 (17)</td>
</tr>
</tbody>
</table>

\(^a\)CRC: colorectal cancer.
\(^b\)HF: heart failure.
\(^c\)GP: general practitioner.
\(^d\)NN: nurse navigator.
\(^e\)N/A: not applicable.

**Recruitment**

Participants were recruited from different settings and through various means. The patients were recruited conveniently [52] from 2 different outpatient clinics at hospital A: 1 HF clinic and 1 gastro-surgical clinic. They were contacted for participation by a designated recruitment nurse during a routine follow-up appointment. They received an information and consent letter from the recruitment nurse and gave their consent to be contacted by the researcher to receive further information and possibly schedule an interview. Of the 20 patients who agreed to be contacted by the research team, only 1 (5%) declined participation after reading the information letter and receiving further information about the study. The RNs were recruited as NNs on personal request by members of the research team or by the head nurse at the department. The GPs were encouraged to participate in the study after receiving general information about it at a meeting for GPs. In addition, they received a reminder by email and as a posting on a web page specifically aimed toward GPs in the area. Those who were willing to participate responded with an email to the researcher and provided their contact information. The researcher contacted the consenting participants and scheduled a suitable time for the interview. The GPs received a gift certificate (value of
approximately €100 (US $108.30) as compensation for the loss of work hours.

Ethics Approval
Ethics approval for the study was obtained from the Norwegian Centre for Research Data (611713). However, ethics approval for the study was considered not notifiable by the Regional Committees for Medical and Health Research Ethics (169884). The research was conducted according to the Declaration of Helsinki, the Regional Committees for Medical and Health Research Ethics, and the research guidelines of the 2 university hospitals.

Informed Consent
The participants were recruited voluntarily and received information about confidentiality, anonymity, and the right to withdraw from the study at any time [53]. Informed consent was obtained after the participants were given information about the nature of the study and aspects of participation. Data were anonymized and securely stored according to Norwegian Centre for Research Data guidelines.

Data Collection
According to the study’s explorative and descriptive design, the aim was to seek new insights into specific issues and serve as a basis for further research [54]. Therefore, this research was conducted with a specific purpose: to inform the design and modeling of an eHealth intervention. This influenced the development of interview guides and data analysis. Semistructured interview guides were developed by the research team and were used to (1) explore the participants’ experiences with the transition phase from hospital to home and the specific needs of the patients during this period and (2) explore their views on the content and functions of a future digital health care solution. At the start of each interview, the patients were asked which digital tools they used daily (eg, smartphone, laptop, and iPad). They were then given a brief overview of the future eHealth solution and possible monitoring devices and asked if such a digital service was something they would be able to operate, either alone or with the help of family members. They were then asked to share their experiences of the period following hospital discharge and their first weeks at home. On the basis of these experiences, they were asked to share what they imagined would be helpful content and functions in a future eHealth posthospitalization follow-up service. The nurses and GPs were also given an overview of the future eHealth service as an introduction to the interviews. The answers from the participants were to form the basis of the content, components, and technical features of the eHealth solution. For a more detailed overview of the questions from the interview guides, please refer toTextbox 1.

The interviews were conducted by the first author both face-to-face and, because of COVID-19 restrictions, digitally through Zoom (Zoom Video Communications) and by phone. A total of 95% (18/19) of the patients were interviewed by phone, and 5% (1/19) were interviewed face-to-face in their home. Of the 6 NNs, 4 (67%) were interviewed face-to-face in an office at their workplace, and 2 (33%) were interviewed via Zoom because of travel restrictions. All the interviews with the GPs (13/13, 100%) were conducted over the phone. During the interviews, the interviewer used follow-up questions such as “What do you mean when you say...?” “Can you elaborate?” and “Is it correct of me to understand what you just said as...?”

All the interviews were audio recorded.
Textbox 1. Examples of questions from the interview guides.

- **Patient questionnaire**—questions to establish digital experience
  - Do you use any digital tools daily (e.g., smartphone, iPad, computer/laptop, or smartwatch/Fitbit)?
  - Can you give me some examples of how and for what you use your digital tools?
  - Do you ever use digital tools in connection with health, disease, or treatment?

- **Patient questionnaire**—questions to help shape the content in an eHealth service
  - During your transition from hospital to home, what did you:
    - experience as problematic?
    - need more information about related to your condition or treatment?
    - need the health care system to help you with regarding managing or complying with the medical regimens you were recommended?
    - need in terms of emotional support?
  - Do you have any thoughts on how the health care system could have offered you support after discharge?

- **Patient questionnaire**—questions to help shape the design and layout of an eHealth service
  - In your opinion:
    - What would be useful components in a postdischarge eHealth service (e.g., illustrations, pictures, type of information, checklists, chat, video, notifications, and reminders)?
    - For an eHealth service to be useful for you in your daily life, what would be important factors to consider?
    - How would you prefer to interact/communicate with health care providers (HCPs) using an eHealth service (e.g., chat, video consultations, or phone)?
  - What is your experience with monitoring devices (e.g., blood pressure, saturation, and weight), and which features seem useful in an eHealth service if you were to assess and monitor your own health condition?

- **HCP questionnaire**—questions to help shape the content of an eHealth service
  - In your opinion/experience:
    - What challenges do patients with heart failure (HF)/patients treated for colorectal cancer (CRC) face after hospital discharge?
    - For patients to cope with long-term illness, what is important to prepare them for?
    - Why and for what reasons do the patients with HF/CRC contact you after hospital discharge?
    - What do you expect from patients after they are discharged from the hospital regarding self-management and adherence?

- **HCP questionnaire**—questions to help shape the design and layout of an eHealth service
  - In your opinion/experience:
    - Where do patients collect information if they have questions regarding their disease and course of treatment?
    - What type of health information is suitable for an eHealth service?
    - What should an eHealth service look like and what seem like core functions and content in such a solution?
  - How do you imagine working with an eHealth service would affect your everyday work?

**Data Analysis**

The audio recordings were transcribed verbatim by the first author and analyzed using a thematic analysis approach in a stepwise process in accordance with Braun and Clarke [55]. In the first step, the transcriptions were read and reread to form an opinion on the overall content and meaning. The fully transcribed interviews were distributed among all the authors, and the texts were subsequently marked and commented on. In the second step, the first author searched for sentences and longer units of text, analyzed them, organized them into possible meaning units, and gave them preliminary codes. In the third step, the meaning units were sorted further and placed into a coding scheme where similar codes were grouped into different categories with a focus on identifying variations, similarities, and differences within each category, aiming to form potential themes. During the third step, all the authors participated by commenting and making suggestions on the coding and categories. In the fourth step, each category was reviewed, refined, and grouped into more distinct subthemes. The
subthemes were identified and named by characterizing content, and by framing differentiated concepts, preliminary main themes were also identified. This back-and-forth process continued until a consensus was reached between all the authors. In the fifth step, categories and subthemes were re-examined, regrouped, renamed, and placed within their correct main themes. In the sixth step, the final schemes were decided on and presented in tables consisting of meaning units, codes, categories, subthemes, and main themes. All the authors participated throughout the various steps of the analysis process to ensure trustworthiness. The coding was performed manually, and no software was used to structure the process.

The data were analyzed groupwise, starting with the transcribed interviews of the patients with HF before starting on the transcribed data material from the patients with CRC. These data were subsequently handled as the previous group, with the various coding schemes systematically compared for similarities, differences, and variations in patient experiences. As the preliminary coding showed similarities across the patient population, the preliminary codes and coding schemes were re-examined, regrouped, renamed, and placed into categories concurrently. Furthermore, as this study aimed to tailor a service to meet the follow-up needs of patients, the findings of the data material from the 2 patient groups formed the basis for the analysis of HCP data material. The stepwise data analysis is shown in Multimedia Appendices 1 to 4.

Results

The findings provided valuable insights into three main themes: (1) expecting information, reassurance, and guidance when using eHealth for HF and CRC self-management; (2) expecting eHealth technology to be comprehensible, supportive, and knowledge promoting; and (3) recognizing both the advantages and disadvantages of eHealth for HF and CRC self-management. For a detailed overview of the main themes and their corresponding subthemes, codes, and data extracts, refer to Multimedia Appendices 1 to Multimedia Appendices 4.

Expecting Information, Reassurance, and Guidance When Using eHealth for HF and CRC

Self-management

The first main theme was supported by 2 subthemes: a need for personalized information and advice about what to expect after discharge and a need for personal interaction to reduce postdischarge uncertainty and anxiety. These 2 subthemes address the supportive care needs of the patients after hospital discharge.

A Need for Personalized Information and Advice About What to Expect After Discharge

The patients with HF described a variety of needs after hospital discharge, and they seemed to have an endless demand for information. Their information needs were mostly related to their diagnosis, the course of the disease, and symptom management. In the period following hospital discharge, many patients with HF described a lack of understanding of what HF was and how the disease would present itself:

I didn’t know I had heart failure. I thought it was a heart attack, not that it was called heart failure. I thought they had fixed me. [Patient with HF 1]

Some patients with HF found it difficult to make individual decisions based on the information they had received, and many lacked confidence in handling their symptoms. Therefore, the patients emphasized that the information they received should be more tailored to fit their individual needs. A patient with HF expressed uncertainty concerning the information he had received:

I used to be allowed to drink 1.5 liters per day, and then they increased it to 2 liters. But if I forget, is that dangerous? Will I start retaining water again? And can I drink more when I exercise and sweat a lot? It would have been nice to know what dangers were associated with it because they say you should stick to what you’re told. [Patient with HF 7]

Many patients with HF described the period after discharge as chaotic. From living a normal life, many were discharged to a life in which they had to pay attention to a disease that they knew little about and take precautions by adjusting to taking several new medications every day. They lacked knowledge about their illness and found it difficult to understand and manage. Moreover, many of the RNs and GPs emphasized that giving information to patients with HF was particularly difficult as the information had to cover a range of different aspects of their lives:

A person with heart failure has so many questions. Some existential, like: Why did this happen to me? But also: How can I live? What can I do? How much can I push myself? Is it dangerous to have sex? Can I go to the store? It is a dramatic and once-in-a-lifetime experience that happens to them, and they have so many questions. [GP 4]

The information needs of patients with CRC were less related to their diagnosis than those of patients with HF. After their tumor was surgically removed, their need for information was mostly dominated by postoperative issues such as bowel function, pain, infection, and leakage, with bowel function causing the most concern. Some of them were also unprepared for the postsurgical pain and the duration of the pain, as described by the following patient:

My bum—it was like barbed wire. It was sown and I had stitches for weeks...and the pain...it lasted for months. I didn’t know it would be like that when they removed my bowel. [Patient with CRC 9]

Other patients with CRC reported that their physical condition returned to normal within the first few weeks following discharge and that their need for information decreased accordingly. Nevertheless, the GPs described unexpected postoperative complications as the main reason why patients surgically treated for CRC made contact after hospital discharge. In addition, some of the patients with CRC needed help with practical matters in the initial weeks spent at home:

Some get complications that may be problematic, but otherwise, they mostly need help with practical things,
Patients treated for CRC also expressed a need to be prepared for what may happen after discharge or “answers to the most common questions that arise after surgery,” as a patient with CRC phrased it. Some also had concerns about nutrition and activity level after returning home from the hospital:

_The information I got from the hospital was that I could eat as normal and move around as much as my body allowed me to, but after the surgery, I couldn’t do as much as I wanted._ [Patient with CRC 2]

Although the 2 patient groups were different in terms of both diagnosis and which symptoms they needed to be aware of after discharge, some challenges were common between them. Most of the patients in both the CRC and HF groups stated that they wished they had been more prepared for how exhausted they would feel after hospital discharge:

_I am a very impatient person, so I wanted to exercise the following day. But everything took longer than I thought, which was very frustrating for me because I thought I could just snap my fingers and all my problems would be solved._ [Patient with HF 6]

_I thought it was fantastic to come home to my family and have them near me. However, I was very exhausted and tired._ [Patient with CRC 7]

The tiredness was described as worrying by the patients, especially as it affected their everyday chores and substantially limited their level of activity. Furthermore, many patients were accustomed to having well-functioning bodies before hospitalization and were not prepared to experience a reduced activity level after discharge. Several of the RNs and GPs recognized activity as an undercommunicated subject and emphasized that both patient groups should be made aware of the importance of restitution after they leave the hospital:

_When they’re discharged, it’s not like they’re expected to be back to their normal selves. The convalescence continues. They must take their time and not wear themselves out because they have a belly that has been opened, and they have to consider the wound. But there’s the housework and the showering and all these everyday things…often small things, but not so small for the patients._ [CRC nurse 1]

### A Need for Personal Interaction to Reduce Postdischarge Uncertainty and Anxiety

The second subtheme, a need for personal interaction to reduce postdischarge uncertainty and anxiety, emerged as a response to the many and various descriptions of the patients’ continuous need for psychosocial support after hospital discharge. Patients with HF described worries and uncertainty about their disease progress, both how long their heart would last and whether it would just suddenly stop. Many described the period after hospital discharge as especially uncertain and frightening, and a lack of information before discharge seemed to contribute to their anxiety and fear of dying:

_I didn’t know what was going to happen. I was constantly afraid. No one called me to ask how I was, and I really missed that because I wasn’t even that old, and I thought I was going to die. Nobody told me anything._ [Patient with HF 4]

Furthermore, some patients with HF were overwhelmed by their many “self-management duties” after discharge. Many also described uncertainty about the future and struggled with existential worries and fear. Patients with CRC also had postdischarge worries and expressed a need to talk to someone after returning home from the hospital. They typically worried about cancer relapse or if the cancer had metastasized so that they would need chemotherapy after the surgery. The waiting period between having the surgery and receiving the histological result was described as particularly straining. In addition, the RNs referred to this as a time of uncertainty and anxiety for the patients:

_I think they are more anxious after the surgery and up until they receive the histology result: that’s when they are scared. And also in regard to further treatment—if they have to do everything all over again or need radiation and chemotherapy._ [CRC nurse 2]

Both patient groups described a need to talk to someone after discharge, and many used family and friends for social support. However, disease-specific issues, symptoms, and advice that included how to conduct necessary changes in their everyday lives were subjects that they wanted to discuss with HCPs:

_I would have liked to ask some questions to someone who knows. That we could have communicated a bit back and forth._ [Patient with CRC 1]

Furthermore, some patients wondered whether their reactions after discharge were normal and expressed a wish to communicate their situation to someone other than their family after they had returned home. In fact, participants from all groups suggested using the digital service to facilitate contact with other patients who had gone through the same thing. Peers were brought up as potential supporters with whom the patients could discuss their various experiences and feelings. Many of the patients sought confirmation from other patients—some sort of affirmation that their reactions and thoughts after discharge did not deviate too much from that of other newly discharged patients. Thus, in this context, peers were suggested as particularly useful supporters:

_Of course, people are different, but there probably are some similarities as well. So, if you could meet others with similar experiences and ask like: what was your reaction to that? Maybe get some kind of confirmation that your thoughts and feelings are the same as everybody else’s._ [Patient with HF 5]

In addition to peers, links to various user organizations, support groups, and validated and reliable websites with scientifically correct information were proposed by both patients and HCPs as something that could potentially support patients.
Expecting eHealth to Be Comprehensible, Supportive, and Knowledge Promoting

The second main theme was supported by the following 2 subthemes: a need for a manageable and useful eHealth solution and a need for communication tools and sources for knowledge acquisition.

A Need for a Manageable and Useful eHealth Solution

Strong agreement existed among all the participants that, if the digital solution was to be manageable and useful, its most important quality should be ease of use:

It has to be as easy to use as a phone. It can’t be difficult to access, then you just wouldn’t be bothered. [Patient with CRC 8]

The participants explained ease of use as easy access, a logical and intuitive interface, clear and visual text, and understandable words and symbols. The solution also had to be beneficial for patients of various age groups:

I think it has to have an easy layout for it to work both for them at age 20 and for those at age 90. It has to be easy, with easy adjustable letters, and a front page with visual and easy things to click on. Not too advanced. [HF nurse 5]

Furthermore, several of the participants in both the patient and HCP groups emphasized a need for reliable and easily accessible information within the solution. However, they thought that the information within the eHealth service should be formulated in a way that everyone could understand, including educated people and those without formal education. The information language should not be too complicated, and the medical terms and formulations should be simplified. One of the GPs accentuated the importance of more straightforward information:

I think a great deal of public information has a high level of learning. However, there should be a point to making the information comprehensible. You really shouldn’t create insecurity, but patients [with HF] need to know why they get breathless. [GP 12]

This GP’s statement was confirmed by participants in both patient groups, stressing that the information within an eHealth service should not cause stress or discomfort. One should also avoid using words that may trigger unnecessary fear. One of the patients with HF was very specific in his advice:

You want it [the medication] to prevent early death. That’s what all the instructions say. But for an anxious person—I don’t think it’s wise that they read the words “early death” because that’s all they’ll see, if you follow? Maybe if it was rephrased to prevent an unfortunate development or bad result. Then they wouldn’t have to read the word death, right? Maybe then they wouldn’t get so anxious. [Patient with HF 1]

Moreover, the HCPs were concerned with making the information of eHealth services explanatory and educational, ideally making the users more knowledgeable and capable of managing their specific disease, including possible precautions and lifestyle changes:

I think it’s important to think educational—to provide them with knowledge they can use long-term. Have I gained weight? Am I breathing more heavily if I walk these steps? And also, it’s wise for them [the patients] to base it on things that are close to them. Like the stairs in their own house or an uphill in their neighborhood. It will make it easier for them to measure. [GP 5]

A Need for Communication Tools and Sources for Knowledge Acquisition

All the participants mentioned several tools and various sources that could promote knowledge and skills among the patients. They explained how regular contact with HCPs after discharge could provide patients with individual and more tailored information that may make them more receptive to changes in their condition and sensitive to the importance of responding to them. The possibility of keeping in touch with the health care services and sending questions and receiving answers from HCPs was suggested by many of the participants, with various ways of contact and communication promoted. Chat was considered to be the fastest and easiest way to connect and communicate. In addition to being time-effective, sending questions through a chat may feel less threatening than reaching out via video or telephone:

I like chat because it is fast, and I feel I can use the time I need to explain. I don’t feel like they are thinking: “she needs to hurry up.” I can take my time and still get answers. [Patient with CRC 8]

Among the HCP population, video was found to be an appropriate and advantageous communication tool as it gave them the possibility to see the person they were talking to. This was confirmed by the patient participants, who emphasized the benefits of relating to a face rather than just to words. Chat was seen as an adequate communication tool for simple and straightforward questions and messages. However, if something needed to be assessed by an HCP, video seemed more trustworthy. By using video, the patient could show their surgical wound or stoma, or their breathing pattern or leg edemas could be assessed by qualified HCPs. One of the GPs also emphasized the following:

I would wish to talk to them. To see them and talk to them. They could show me things, like swollen ankles or something, and also I can get an impression how they breathe. Or if they’ve had a bowel operation...their wound or skin. You do get a better impression with video. [GP 11]

Many participants in both the patient and HCP populations also suggested answering questions regularly, such as questionnaires or checklists, as something that may promote appropriate self-management activities and keep patients updated on their condition. Some of the HCP participants suggested that answering questions within the eHealth service could function as a reminder for the patients—regular cues that reminded the patients “to do what they’re supposed to do,” as one of the RNs phrased it. By logging on to a digital system and actively replying to questions regularly, patients, especially those with HF, thought that they would become more aware of their
behavior and habits, as well as becoming more receptive and willing to engage in appropriate and health-promoting self-management activities. The idea of receiving feedback on checklists was also emphasized as particularly beneficial:

Checklists would be great, and blood pressure, follow-up regarding medication and maybe also weight, like: Have you gained weight? How is your weight? Have you retained water in your body? Do you have to increase your medication? Those are the sorts of things where you don’t know what to do, and then you could write like: I am feeling like this and that—what shall I do? [Patient with HF 9]

However, it was emphasized by one of the GPs that the questions in the eHealth service had to be disease- and symptom-specific so it would be easy for the patient to connect it to their specific condition:

Take heart failure, for instance, if there was something you wanted to measure, you could ask: How many stairs can you climb? How many meters can you walk on a flat road? Your morning weight? That will give them something to compare and they can see changes. I’m very skeptical to “how are you” questions because that’s very subjective and not necessarily related to the condition. [GP 5]

During the interviews, the participants were asked to share their thoughts and experiences regarding home monitoring and vital signs. This was a subject on which the participants had different opinions. Most participants, both patients and HCPs, had a positive attitude toward home monitoring. They proposed that, if the patients monitored their vitals at home after receiving proper training, it could help them gain a better overview of the disease progress and make them more attentive to symptoms and complications and be more in control of their health condition:

For most patients, I think it would feel very safe and reassuring to know that they have something concrete to pay attention to. I think it could be meaningful for them during the first period. I also think that saturation, weight, and blood pressure are familiar for most patients today. It may also give them more understanding and insight into their own disease. [HF nurse 6]

Some of the participants also stated that home monitoring could form the basis for information and reduce the number of visits to the physician’s office. However, some of them were skeptical about home monitoring and claimed that leaving patients in charge of such measurements might be perceived as burdensome and potentially cause unnecessary worries for the patients, especially if the measurements showed discrepancies:

You can get a bit caught up in it [home monitoring] as well, and when you have gone through something like this, you’ll probably be monitored pretty good anyway, so I don’t know if it is such a good idea. [Patient with CRC 5]

Nevertheless, although not every patient or GP saw the benefits of home monitoring, some patients with HF were used to taking various measurements, such as measuring their blood pressure, regularly counting their heart rate, or paying attention to their weight. Many of these patients, along with some of the GPs, proposed that it would be beneficial if the digital solution had graphical or statistical visualizations of the various measurements that the patients had taken along with feedback on the measurements if they were irregular:

I am keeping an eye on my weight, so maybe if I had the possibility to enter the numbers and see them as a graph. I think that would be interesting. [Patient with HF 2]

Recognizing Both Advantages and Disadvantages of eHealth Services for NCD Self-management

The last main theme comprised the following 2 subthemes: recognizing eHealth as a tool for follow-up care and concerns about eHealth as a tool for follow-up care.

Recognizing eHealth as a Tool for Follow-up Care

All the RNs were positive toward eHealth and argued that digital follow-up care would prolong the period in which patients were under supervision from health care services, which could strengthen the relationship between the patient and the hospital. However, they primarily argued that a digital follow-up service could be supportive for patients during the vulnerable phase following discharge:

I think it [an eHealth service] may serve as a connector between the patients and the hospital after discharge. The first few days after returning home are the most uncertain, so I think that every patient may benefit from being watched over by someone from the health care services who checks if they manage everyday life at home. And then they can rest knowing that they’re not all by themselves. [CRC nurse 1]

RNs also emphasized that an extended follow-up period, which an eHealth service may provide, could offer the patients more adapted information. This could lead to an increased sense of security and make the patients and families more capable of managing life at home. In addition, a digital solution with symptom registration could lead to the early detection of changes and subsequently prevent readmissions. Patients with CRC also recognized potential benefits of digital follow-up care, especially in connection with postsurgical complications, as described by the following participant:

I think a digital solution to help people post discharge would be helpful for those with complications. If there was something wrong with the surgical wound for instance. [Patient with CRC 2]

A positive attitude toward digital follow-up care was supported by most patients with HF. An eHealth service could be reassuring for patients after discharge and lower the threshold for asking questions. Furthermore, several of the patients with HF shared stories about how they felt insecure, lonely, or “left to themselves” after they came home from the hospital, and some were under the impression that a digital solution could
have reduced some of the negative emotions they experienced after discharge:

*When I got home after discharge, the house was freezing cold, and I was all alone. I felt really lonely. Coming home to a cold and empty house, without anybody around you...That’s what I remember as the worst part. So maybe if I had an iPad? Or access to a chat or something. Maybe I wouldn’t have felt so completely left alone.* [Patient with HF 5]

The GPs mostly viewed digitalization in health care as beneficial as a nurse-assisted eHealth service could potentially make health care services more approachable by simplifying communication and lowering the barrier to contact. They also suggested that maintaining contact through a digital solution may feel less threatening for the patients, as well as putting less strain on the health care system. This was a view shared by many of the nurses, especially the CRC nurses.

**Concerns About eHealth as a Tool for Follow-up Care**

Some of the GPs expressed concerns about this type of follow-up care. They seemed worried that a nurse-assisted eHealth service would disturb the patient-GP relationship. They recognized that the eHealth service could be a tool to help patients cope and give them answers to many of the questions they had after discharge, but a digital solution should never interfere with the interaction between them and their patients, as the following GP emphasized:

*I think they need to be reminded to contact their GP so that they can get help to assess the situation or control things, because I think there are quite a few readmissions. So, it [eHealth] could be a smart way to reach people when they are in trouble and need help, but I don’t think it’s wise to let it replace the GPs’ evaluations.* [GP 9]

Some of the patients, mainly the patients with CRC, expressed skepticism about the need for a digital follow-up service. In total, 2 factors were highlighted as particularly challenging when it came to digitalization of the health care system, with the first being the human factor. Communicating and receiving follow-up care without physical contact or connection with an actual person was viewed by some of the patients as foreign and “cold.” One of the patients with CRC said the following:

*Isn’t that just a complete waste? In my sense, it is much better to have contact with people over the phone or with your GP. You lose all contact. It is just a machine.* [Patient CRC 4]

However, the digital competence factor seemed to cause more concern for other patient participants. They indicated that eHealth and its technical features would be difficult for some people to understand, and they questioned whether everyone would have sufficient digital competence to operate the solution:

*I think this digital solution is very appropriate. But I’m not sure that everyone will find it convenient to use. Some will not have the skills, and some will have a bit of an aversion to this computer world. Not everyone can use this type of equipment.* [Patient with CRC 6]

A third issue regarding eHealth was highlighted by some of the patients with CRC: who would benefit from using an eHealth service? The patients with CRC seemed to believe that digital follow-up care was most appropriate for patients who experienced some type of surgical or medical complication. Many of the patients with CRC used the phrase “differently sick than me” to describe patients who would benefit from using an eHealth service after discharge. A patient with CRC said the following:

*I think if I was different. Say I had metastasis. Then I would want to have contact, but as long as I felt well and they said that there wasn’t anything wrong...then I just would have wanted to go to my regular follow-ups. I think it would have been more burdensome if I had an app and felt that I had to write to someone. That would have taken up too much of my time.* [Patient with CRC 6]

“Differently sick” included everything from having surgical complications to having a stoma or being diagnosed with metastasis. The prevailing view of the patients with CRC was that, if the operation and postoperative course went without complications, there was no need for digital follow-up care. Having to deal with a digital solution after discharge was thought to add to the treatment burden rather than decrease it. These patients’ view of digital follow-up care stands in contrast to that of the CRC nurses, who claimed that newly discharged patients frequently contacted the hospital ward with various questions after CRC surgery. Answering phone calls from insecure and worried patients was described as work and resource-demanding. The CRC nurses spoke about how they expected the patients to understand the information they were given during hospitalization. In addition, they provided the patients with a discharge letter and expected them to collect the necessary information from there or call their GP with any additional questions, as described by the following nurse:

*We often experience that patients call the hospital ward because they are insecure after discharge. Even though the discharge letter clearly says they should contact their GP. We get quite a few phone calls.* [CRC nurse 2]

**Discussion**

**Principal Findings**

This study applied a qualitative interview approach to explore various stakeholders’ perspectives on self-management needs after hospital discharge and investigate how a future eHealth service can be best designed to foster patient acceptance, support self-management, and ease the transition from hospital to home. We found that patients with both HF and CRC had unanswered questions and faced various challenges after hospital discharge. Some struggled to understand which self-management tasks were necessary and what precautions they should take to avoid complications or exacerbations. The statements from the patients regarding posthospitalization self-management challenges were confirmed and expanded upon by the HCPs. In addition, the
participants shared many valuable opinions and ideas about the content and functions of a future eHealth service.

The first and overarching main theme demonstrated how additional information and follow-up care are necessary for patients during the transition to home after hospital discharge regardless of diagnosis. The patients in this study described common challenges in their daily lives, including fatigue; confusion regarding activity level; and psychosocial challenges such as negative thoughts, worries about the future, and a general need for more support. These findings are supported by existing literature describing challenges following discharge for patients with both cancer [22] and HF [56]. In addition to information, the patients in this study emphasized a need for more tailored advice about what to expect after discharge and personal interaction to reduce postdischarge uncertainty and anxiety. Tailored information and personal interaction are conditions that may be closely intertwined and should be seen in relation to each other as insufficient information or a lack of advice about disease management may lead to extensive worrying and a lack of confidence to engage in necessary self-management after hospital discharge [19]. Moreover, depression and anxiety may impede an individual’s ability to engage in self-management behaviors [57]. For many patients, especially those living with long-term illnesses, hospital discharge often marks the start of a new round of self-management activities [58]. Thus, the findings from this study highlight the importance of “equipping” patients with NCDs with more tailored knowledge and skills to reduce or prevent psychological conditions that may hinder self-management.

The second main theme captured the participants’ views and ideas on the design and technical functions of a digital solution as well as identifying relevant content that could meet the support needs of patients with HF and CRC during the transition phase from hospital to home. As the patient participants and GPs in this study described common challenges across the 2 patient groups, there seem to be various core functions that could shape the content of a digital platform. Easily accessible, understandable, and nonfrightening disease-specific information; multifaceted knowledge-enhancing functionalities; and different communication sources such as chat, video, checklists, and home-monitoring devices were the most prominent features suggested by the participants. According to Nymberg [59], many patients can see possibilities with the use of eHealth as an improvement, alternative, or complement to existing health care. However, there is a strong need for user-friendly and well-adjusted digital tools compatible with patients’ needs [31,59].

An important finding of this study is that being able to exchange messages and receiving informational support from HCPs would help reinforce the self-management skills of patients with NCDs and give them a better understanding of their medical condition with its accompanying symptoms and complications. In addition, receiving emotional support from HCPs after discharge was thought to help patients cope with their postdischarge worries and the need for practical advice. Evidence exists for the positive effects of eHealth on patients’ perceived support [39]. Support is essential to help individuals accomplish self-management tasks, and it is an important strategy to reduce the burden of chronic disease [60]. Moreover, from a patient perspective, acceptance of technology is greater when it is not perceived as replacing in-person care [61]. Therefore, it is important that eHealth services have a “human component” and serve as a complement to, not a replacement for, usual care [16,59]. In that sense, nurses may be an asset in future eHealth solutions for patients with NCDs as instigators of contact with them after hospital discharge. Regular contact and follow-up care from designated nurses may also contribute to fostering patient adherence to treatment [62].

The third main theme identified the participants’ views on eHealth in general and specifically on eHealth as a tool for self-management. The participants in this study had different opinions on the value of and need for eHealth. Although most participants in both the patient and HCP populations seemed curious and positive toward eHealth, some expressed skepticism. This is in line with other research showing that many patients have different perceptions and expectations of eHealth [63,64]. On the one hand, eHealth may be viewed as something difficult and troublesome, and on the other, it may be seen as something that makes things easier. In this study, the various perceptions of eHealth and digital follow-up care seemed to be related to human or technological factors. Participants most in favor of eHealth attributed this to the advantages of patients being able to contact and communicate with HCPs at the time of exacerbations or worries, thus receiving follow-up care when it is perceived as most useful and needed. Moreover, having access to information within an eHealth solution and being able to repeat and confer this information with HCPs were also considered major benefits. Participants who were less positive toward eHealth and digital follow-up care seemed to worry about the “faceless” interaction within an eHealth service and that digitalization would disturb the personal relationship between the patient and HCP. They also feared that valuable information and time would be lost if they were to communicate digitally during an exacerbation. Not being able to manage the various technological aspects of eHealth was also a source of concern, a notion highlighted by several of the participants from all groups.

Overall, the findings from this study suggest that the digitalization of health care provides opportunities and challenges. It seems that patients’ expected benefits of using eHealth might be seen as an important predictor of their willingness to use it. A future eHealth service for patients with HF or patients surgically treated for CRC in the transition from hospital to home could potentially reduce the treatment burden for some as it may support self-management strategies and decrease the number of appointments and personal visits within the health care system. It may also enhance patient’s knowledge and understanding of their condition and provide them with a sense of control. However, to foster patient acceptance, it seems equally important that a future eHealth solution have a human component and focus on becoming a positive contribution to the patients’ daily life and not just on the negative aspects of living with a long-term illness. As this study indicates, some patients may not perceive it as useful to be reminded regularly about having a chronic illness, especially those who already

https://humanfactors.jmir.org/2023/1/e39391

JMIR HUMAN FACTORS

Wathne et al

JMIR Hum Factors 2023 | vol. 10 | e39391 | p.506

(page number not for citation purposes)
have a social system that provides them with sufficient knowledge and support. Furthermore, some of the time saved by using eHealth and, thus, not having to physically attend health care appointments will be substituted by additional self-monitoring work and other health care tasks [65] such as taking various measurements, answering checklists, or digitally communicating with HCPs. For some patients, this may be perceived as adding to the treatment burden [65].

Comparison With Prior Work

Patients value education on disease and disease management, specifically information about health status and symptoms, exacerbations, and new challenges [62,66]. “The more you know, the safer you feel” has been expressed by patients with other chronic diseases [67]. Nevertheless, some might struggle with transforming the information they receive during hospitalization into action after discharge. In addition, the healthcare system today seems to shift a steadily growing list of self-management responsibilities and tasks to the patients’ posthospital discharge, which requires considerable effort from the patients [3]. Hence, patients need self-management support to respond to physical and mental changes and manage their day-to-day challenges and decisions after hospital discharge [13]. Self-management integration is an ongoing process that includes various phases. Seeking effective self-management strategies and creating routines and plans of action are highlighted as 2 crucial steps [14]. However, as the findings of this study demonstrate, many participants described it as challenging to independently seek appropriate strategies and create proper routines in everyday life shortly after hospital discharge. Some lacked a basic understanding of their diagnosis and its symptoms or unexpected complications they should be aware of. Hence, to better meet the supportive care needs of patients with chronic conditions and help them with their self-management tasks, it could be beneficial to provide them with an extended support system through an eHealth service that offers them information, practical advice, and psychosocial support after hospital discharge.

The importance of engaging in self-management activities after discharge and developing more tailored eHealth solutions has been promoted in earlier research [24,31,68]. The term “perceived usefulness” is an important predictor of the acceptance of eHealth, and an eHealth service is more likely to be accepted if the perceived benefits of using the service are outweighed by the negative consequences of having to act on and deal with the disease [63,69]. This study suggests that each patient group had different needs regarding self-management support after hospital discharge. However, the findings also showed common self-management challenges after care transitions across the patient groups. Thus, developing a generic intervention that “fits all” may be possible assuming that the service contains targeted information and functions tailored to fit each diagnosis and self-management support needs. Patients’ perceived benefits of using eHealth could also increase if the service is developed in response to users’ needs rather than as a technological innovation [37]. Moreover, research shows that patients with NCDs want professional support through eHealth services, including human contact to help them address health issues [63]. However, eHealth cannot substitute the personal interaction between patients and HCPs. A nurse-assisted eHealth service will allow patients to communicate their self-management challenges and receive self-management support from a designated NN through a digital service. By designing an eHealth service that considers the holistic needs of patients, clinicians (ie, NNS) can support patients in their transition to self-management [13]. Continuous self-management support from HCPs after hospital discharge could also help patients become more knowledgeable and, at the same time, make them more confident in their skills to manage their illness [60,70]. This could increase patients’ compliance with their healthcare regimen, which can lead to a reduced number of hospital admissions [71]. As the risk of rehospitalization is high during the first weeks at home (30% for patients with HF [72] and 15% for patients treated for CRC) within the first 30 days after discharge [73], the transition period from hospital to home seems to be an appropriate time to offer digital follow-up care to patients with long-term illnesses such as HF or CRC.

Strengths and Limitations

This study has several strengths. It included different stakeholders’ views and opinions, which gave varied insights into self-management challenges after hospitalization. Moreover, the study participants varied in gender, age, and educational level, which may have provided this study with a broad perspective on how a nurse-assisted eHealth service could be best designed. Furthermore, in this qualitative study, the aspects of trustworthiness were covered by establishing credibility, dependability, confirmability, and transferability [74]. The credibility and dependability were assured by describing the analytical process in detail and using researcher triangulation throughout the analytical process. Confirmability was assured by presenting the various steps of the analysis, along with a broad overview of data extracts from the participants, in an appendix to make it possible for the reader to agree with and understand the logic of the findings. Transferability was assured by providing the reader with a detailed description of the background and context of the study and focusing on the participants’ stories when presenting the analysis [74].

This study also has some limitations. First, the participants answered questions about an imaginary digital solution. Thus, some of the perspectives and suggestions from the participants will be difficult to transfer and apply within the limits of the future service. Second, the patients in this study answered questions about postdischarge supportive care needs retrospectively, which may have introduced a memory or recall bias [75]. Third, the RNs were recruited for this project because of their interest in eHealth and motivation to provide follow-up care after hospital discharge. Furthermore, the GPs who volunteered to participate may have been more engaged, motivated, and interested in the use of eHealth than the average GP. This may have affected the transferability to both nurses’ and GPs’ perceptions of eHealth in general. Finally, most of the recruited participants who were surgically treated for CRC experienced their discharge period as relatively complication-free. Therefore, their perspectives may not be applicable to the general population of patients with CRC, specifically to patients who experienced complications. Perhaps
Conclusions

This study explored stakeholders’ experiences with supportive care needs and their perspectives on eHealth as transitional care for patients with HF and patients surgically treated for CRC as part of an iterative development process of a planned nurse-assisted eHealth service. Both patient populations need specific and tailor-made information on what to expect when transitioning from hospital to home to be as well prepared as possible for self-management tasks. Moreover, they need guidance on how to monitor their health conditions and options for communicating changes to HCPs to avoid uncertainty and anxiety. At the same time, the results indicate that eHealth follow-up services must be adapted according to the severity of the patient’s condition and level of self-management confidence.

This study is valuable as it contributes necessary information from both primary (ie, patients) and secondary (ie, HCP) sources that can ensure the relevant and safe follow-up of patients with NCDs during challenging phases of a care pathway. It suggests eHealth as a possible asset with the potential to bridge the health care void experienced by many patients following a hospital admission. It may fill the resource and knowledge gaps faced by patients with NCDs when performing self-management tasks and prevent unnecessary anxiety and uncertainty among patients. Furthermore, this study stresses the need to tailor the content, functions, and delivery mode of eHealth services to achieve a patient-centered, feasible, and acceptable follow-up after hospitalization. In addition, it may add value to the planning and development of eHealth interventions for other patients with NCDs.

Acknowledgments

The authors would like to thank the participants of this study for sharing their experiences and for their valuable ideas on the design and content of an eHealth service.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Analysis steps of interviews with patients with heart failure.
[DOCX File, 28 KB - humanfactors_v10i1e39391_app1.docx]

Multimedia Appendix 2
Analysis steps of interviews with patients with colorectal cancer.
[DOCX File, 26 KB - humanfactors_v10i1e39391_app2.docx]

Multimedia Appendix 3
Analysis steps of nursing interviews.
[DOCX File, 27 KB - humanfactors_v10i1e39391_app3.docx]

Multimedia Appendix 4
Analysis steps of general practitioner interviews.
[DOCX File, 27 KB - humanfactors_v10i1e39391_app4.docx]

References


Abbreviations

CRC: colorectal cancer
GP: general practitioner
HCP: health care professional
HF: heart failure
NCD: noncommunicable disease
NN: nurse navigator
RN: registered nurse
Exploring the Cross-cultural Acceptability of Digital Tools for Pain Self-reporting: Qualitative Study

Syed Mustafa Ali1,2, MPH, MSc; Rebecca R Lee1, PhD; John McBeth1, PhD; Ben James3, BA; Sean McAlister3, BA; Alessandro Chiarotto4, PhD; William G Dixon1, PhD; Sabine N van der Veer2, PhD

1Centre for Epidemiology Versus Arthritis, Division of Musculoskeletal and Dermatological Sciences, Manchester Academic Health Science Centre, The University of Manchester, Manchester, United Kingdom
2Centre for Health Informatics, Division of Informatics, Imaging and Data Science, Manchester Academic Health Science Centre, The University of Manchester, Manchester, United Kingdom
3uMotif, London, United Kingdom
4Department of General Practice, Erasmus University Medical Center, Rotterdam, Netherlands

Corresponding Author:
Syed Mustafa Ali, MPH, MSc
Centre for Epidemiology Versus Arthritis, Division of Musculoskeletal and Dermatological Sciences, Manchester Academic Health Science Centre
The University of Manchester
Oxford Road
Manchester, M13 9PT
United Kingdom
Phone: 44 161 275 1644
Email: syedmustafa.ali@manchester.ac.uk

Abstract

Background: Culture and ethnicity influence how people communicate about their pain. This makes it challenging to develop pain self-report tools that are acceptable across ethnic groups.

Objective: We aimed to inform the development of cross-culturally acceptable digital pain self-report tools by better understanding the similarities and differences between ethnic groups in pain experiences and self-reporting needs.

Methods: Three web-based workshops consisting of a focus group and a user requirement exercise with people who self-identified as being of Black African (n=6), South Asian (n=10), or White British (n=7) ethnicity were conducted.

Results: Across ethnic groups, participants shared similar lived experiences and challenges in communicating their pain to health care professionals. However, there were differences in beliefs about the causes of pain, attitudes toward pain medication, and experiences of how stigma and gender norms influenced pain-reporting behavior. Despite these differences, they agreed on important aspects for pain self-report, but participants from non-White backgrounds had additional language requirements such as culturally appropriate pain terminologies to reduce self-reporting barriers.

Conclusions: To improve the cross-cultural acceptability and equity of digital pain self-report tools, future developments should address the differences among ethnic groups on pain perceptions and beliefs, factors influencing pain reporting behavior, and language requirements.

Introduction

Pain Inequalities

Chronic pain affects approximately 28 million people in the United Kingdom alone [1], causing both personal and economic burden [2]. To reduce this burden, it is essential to accurately measure pain, know its causes, and estimate its impact on people’s lives [3]. There are inequalities in pain prevalence, pain intensity, and pain treatment that have been linked to people’s characteristics, including their socioeconomic status, geographical location, and ethnicity [4,5]. For example, lower socioeconomic status is associated with higher bodily pain levels...
In the United Kingdom, Australia, and Germany, particularly in older people [6], and pain is more prevalent among the Black and Asian ethnic minorities [7]. Asians are less likely to receive pain medication than White patients [8], and Black individuals may have different pain management preferences and expectations [9].

**Influence of Culture and Ethnic Background on Pain Experience**

Inequalities in pain may be partly explained by the influence of culture and ethnicity on pain perception and reporting. A person’s cultural and ethnic background may affect the way he/she perceives, experiences, and communicates pain [10], and people from different ethnic groups tend to give different meanings to pain [11]. In turn, these inequalities may impact the quality and content of patient-provider communication on pain [12,13].

The influence of the cultural and ethnic background on an individual’s pain experiences and reporting behaviors makes it challenging to develop tools for self-reporting pain that are acceptable and valid across ethnic groups. For example, a review of the cultural adaptations of the Pain Catastrophizing Scale found that construct (ie, varying correlation with other pain scores) and structural (ie, differences in subscales) validity varied across translated versions [14]. Moreover, a review by Booker and Herr [15] found that many pain assessment tools lacked evidence of their validity and reliability in ethnically diverse populations. Another review reported that digital pain self-report and self-management apps seldom offered culturally tailored aspects [16], potentially hampering their cross-cultural acceptability. Similarly, a review of smartphone-based pain manikins found that the manikin appearance could seldom be culturally personalized [17].

**Objectives of This Study**

The aim of this study was to inform the design and development of cross-culturally acceptable digital pain self-report tools by better understanding individuals’ pain experiences and reporting behaviors across ethnic backgrounds. The specific objectives were to explore similarities and differences across ethnic groups in (1) the description of pain experience and its reporting and (2) user requirements for digital pain self-report tools by using a smartphone-based pain manikin as an example. We expect this to contribute to acceptable and, ultimately, valid digital pain self-report for people living with a painful condition, irrespective of their ethnic backgrounds.

**Methods**

**Study Design**

We conducted 3 web-based workshops, each consisting of a focus group discussion and a user requirement exercise. The focus group discussions addressed the first objective. This phenomenological approach acknowledges and explores the subjective experience, which can be used to develop or reorient our understanding of the phenomenon under consideration [18]. We explored the phenomenon of pain experience, its reporting, and how it is embedded within individuals’ cultural and ethnic backgrounds. We used the consolidated criteria for reporting qualitative research checklist to guide reporting of this part of our study [19]. For the second objective, we analyzed user requirements by using the Table of Specifications approach [20] to guide discussions on important aspects of digital pain self-report tools by using a smartphone-based manikin as an example. This approach attempts to translate a set of concepts (in our case, aspects of pain experience and reporting) into a set of items that can be used to assess them.

**Ethics Approval**

The study received a favorable opinion and Health Research Authority approval from the National Health Services Westminster Research Ethics Committee (ref 21/PR/0342).

**Eligibility and Recruitment of Participants**

Adults (older than 18 years) were eligible to take part in this study if they lived in the United Kingdom and self-identified as (1) living with a primary (ie, pain without any underlying condition) or secondary pain condition (eg, ankylosing spondylitis, rheumatoid arthritis) for more than 3 months and (2) being Pakistani or Bangladeshi, Black African, or White British. Using a purposive sampling approach, we invited people of specific ethnicities who had participated in a related study on the feasibility of a pain self-reporting tool using a smartphone-based pain manikin (Ali SM et al, unpublished data, January 2023). We also recruited potential participants via online community groups (eg, WhatsApp groups for Black Africans, a Facebook group for Pakistanis), as well as online groups of people with an interest to take part in research studies through convenient sampling. We shared a study flyer (Multimedia Appendix 1) with them, after which people could express their interest in taking part. One researcher (SMA) then determined people’s eligibility by telephone screening and asked those eligible to provide informed written consent via email.

**Data Collection**

We organized 3 web-based ethnicity-specific workshops consisting of focus groups followed by a user requirement exercise on Zoom: one with South Asians, one with Black Africans, and one with White British. All workshops had the same topic guide (Multimedia Appendix 2), and each was scheduled to last for 2 hours. Before the workshops, participants completed a web-based questionnaire, capturing key demographics (age, gender, ethnicity, employment status) and questions related to their pain experience [21] and perception and beliefs [22]. We assigned a 4-digit code to all consenting participants and followed established institutional guidelines to ensure confidentiality of their data. They also received workshop details via email and were offered support with joining the web-based workshop, if needed. Two researchers (SMA and SNvdV) facilitated the workshops and presented the ground rules for the session at the start of the workshop (eg, providing a safe space for sharing opposing opinions, keeping discussions private within the group). Representatives from uMotif Limited (BJ and SMA), our technology partner, developed and presented the mock screens for feedback but they were not involved in other aspects of the data collection or in the data analysis.
Cultural Similarities and Differences in Pain Experience and Its Reporting

The topic guide for the focus group discussions on pain experience and reporting and its relationship with culture (objective 1) was informed by the literature [21,23-25] and included topics such as pain experience, pain perception, pain report and communication, and pain assessment. Focus group discussions were audio-recorded and transcribed verbatim. Once the transcriptions were ready, we anonymized the transcripts and destroyed the audio recordings.

User Requirements for Digital Pain Self-report Tools

To prompt discussions on important aspects of pain self-reporting (objective 2), we demonstrated the Manchester Digital Pain Manikin app [26]—developed by uMotif Limited—as an example of a digital pain self-report tool (see Figure 1). People can use the app to report overall pain intensity on a numeric rating scale from 0 to 10, location-specific pain intensity on a 2D gender-neutral body manikin, and a free text pain diary to elaborate on the manikin drawing. After the demonstration, focus group participants were split into 2 smaller breakout groups to discuss user requirements, including what they would want to report about their pain (ie, pain aspects) and how (ie, app features) and why they considered these aspects and features important (see Table 1). Digital pain self-report tools can have multiple purposes (including supporting self-management, guiding clinical decisions, collecting data for research), and we did not specify any particular purpose at the start of these discussions. The facilitators recorded the breakout groups’ responses in a shared Google doc.

Figure 1. Screenshots of the Manchester Digital Pain Manikin app (developed by uMotif Limited, copyright University of Manchester and uMotif, 2020), which we used as an example of a digital pain self-report tool. A. Numeric rating scale for overall pain intensity; B. Front view of the body manikin with pain drawing; C. Back view of the body manikin with pain drawing; D. Pain diary.

Table 1. Questions to guide breakout group discussions on user requirements for the pain self-report tool.

<table>
<thead>
<tr>
<th>Question type</th>
<th>Question on user requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>What (Q1)</td>
<td>What pain aspect or app feature would be important for you to report?</td>
</tr>
<tr>
<td>Why (Q2)</td>
<td>Why is that aspect or feature important?</td>
</tr>
<tr>
<td>How (Q3)</td>
<td>Do you feel the aspect or feature is currently available in the Manchester Digital Pain Manikin app?</td>
</tr>
</tbody>
</table>

Data Analysis and Synthesis

Cultural Similarities and Differences in Pain Experience and Its Reporting

To analyze the transcripts of the focus group discussions (objective 1), we utilized an interpretive analysis approach, also referred to as hermeneutic phenomenology [27]. This approach has been used in previous studies to understand the lived experience of pain [28]. Two researchers (SMA and RRL) reviewed the transcripts line-by-line independently to immerse themselves in the data; both had experience of qualitative data collection and analyses in the fields of public health and health psychology, respectively. One researcher (SMA) assigned codes to all relevant statements to find patterns, linked them across transcripts, and discussed these in the context of the participants’ cultural background with the other researcher (RRL). Both researchers used their own cultural background to interpret textual data, codes, and themes, which emerged from the data. SMA recorded all the emerging codes in a codebook alongside illustrative quotes, iteratively refining the codebook after reviewing each transcript and discussing them with RRL. Once the codebook was finalized, SMA reapplied it to all the transcripts and drew themes to ensure consistency. Under each theme, we first synthesized similarities in people’s pain experience across ethnic groups and then highlighted differences in their viewpoints that may be linked to their ethnic
backgrounds. We managed all qualitative data (ie, codebook, illustrative quotes) by using Microsoft Excel.

**User Requirements for Digital Pain Self-report Tools**

User requirements for pain self-report tools related to important pain aspects and app features (objective 2) were thematically synthesized by 2 researchers (SMA and SNvdV) to identify similarities and differences between ethnic groups. We then invited participants from across ethnicity-specific workshops to attend another web-based workshop. People could express their interest via email and were offered a place on first come, first served basis, while ensuring a balanced representation across ethnic groups. The aim of the workshop was to check for accuracy of our findings and whether these resonated with participants’ experiences and preferences (ie, member-checking exercise). For this, we asked participants for feedback on our synthesis of the user requirements as well as on mock-ups for the Manchester Digital Pain Manikin app to illustrate how some of the identified key requirements for pain self-report tool could be translated into app functionalities.

**Results**

**Participants’ Characteristics**

In total, 23 adults (14 females, 61%) took part across the workshops. Table 2 shows that the majority of the participants (13/23, 56%) were aged 45 years and older, did not have English as their native language (12/23, 52%), and had experienced pain for 4 years or more (15/23, 65%). Regarding participants’ pain perception and beliefs, all thought that pain intensity varied but was always present, and 14 (61%) participants felt that they did not know enough about their pain.
Table 2. Characteristics of the study participants (N=23).

<table>
<thead>
<tr>
<th>Characteristics, response categories</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td></td>
</tr>
<tr>
<td>25-34</td>
<td>5 (22)</td>
</tr>
<tr>
<td>35-44</td>
<td>5 (22)</td>
</tr>
<tr>
<td>45-54</td>
<td>4 (17)</td>
</tr>
<tr>
<td>55-64</td>
<td>7 (30)</td>
</tr>
<tr>
<td>65+</td>
<td>2 (9)</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>9 (39)</td>
</tr>
<tr>
<td>Female</td>
<td>14 (61)</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
</tr>
<tr>
<td>South Asian(^a)</td>
<td>10 (45)</td>
</tr>
<tr>
<td>Black African</td>
<td>6 (27)</td>
</tr>
<tr>
<td>White British</td>
<td>7 (32)</td>
</tr>
<tr>
<td>Employed</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>10 (45)</td>
</tr>
<tr>
<td>No</td>
<td>13 (55)</td>
</tr>
<tr>
<td>Is English your native language?</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>11 (48)</td>
</tr>
<tr>
<td>No</td>
<td>12 (52)</td>
</tr>
<tr>
<td>How long have you been experiencing pain? (years)</td>
<td></td>
</tr>
<tr>
<td>≤3</td>
<td>8 (35)</td>
</tr>
<tr>
<td>4-10</td>
<td>6 (26)</td>
</tr>
<tr>
<td>&gt;10</td>
<td>9 (39)</td>
</tr>
<tr>
<td>My pain varies in intensity but is always present</td>
<td></td>
</tr>
<tr>
<td>Agree</td>
<td>23 (100)</td>
</tr>
<tr>
<td>Disagree</td>
<td>0 (0)</td>
</tr>
<tr>
<td>I do not know enough about my pain</td>
<td></td>
</tr>
<tr>
<td>Agree</td>
<td>14 (61)</td>
</tr>
<tr>
<td>Disagree</td>
<td>9 (39)</td>
</tr>
<tr>
<td>If I am in pain, it is my own fault</td>
<td></td>
</tr>
<tr>
<td>Agree</td>
<td>5 (22)</td>
</tr>
<tr>
<td>Disagree</td>
<td>18 (78)</td>
</tr>
</tbody>
</table>

\(^a\)Included people from Pakistani and Bangladeshi backgrounds.

Cultural Similarities and Differences in Pain Experience and Pain Reporting

Participants across all ethnic groups indicated that their culture influenced how they perceived pain (eg, what causes pain), how they managed it (eg, whether to take medication), and how they communicated about their pain and with whom. Four main themes emerged from our interpretive analysis, namely, perceived causes of pain, approaches and attitudes to self-treatment and management, frustration and embarrassment when communicating about pain with others, and lack of experience with formal pain assessment tools. Below, we describe each theme in more detail alongside selected illustrative quotes, with additional quotes supplied in Multimedia Appendix 3.

Theme 1: Perceived Causes of Pain

Most participants described their pain experience as agonizing and explained how it was to live with pain, what caused their pain in their perception, and what impact it had on them. Across ethnicities, participants described their pain in similar ways, including that they were always in pain but that it fluctuated.
They referred to good days or bad days when pain was less or more, respectively. Female participants talked about gender norms such as caring responsibilities and domestic chores as an inevitable cause of their pain.

...Females do all the domestic chores and with time and age it's bound to happen. These things are supposed to kick in and women do complain...so complaining about pain is just the norm for women I think. [South Asian female]

Only participants from South Asian background perceived food type to be a cause of their pain.

...Food that has a lot of spice, perhaps has a lot of oil, and you use ghee based substances, which can cause a greater reaction in my opinion. [South Asian male]

South Asian and White British participants also perceived weather conditions to be a potential cause.

...It could be the weather here, because when I go to...like I've been to Spain, I've been to Pakistan, Dubai, it's very hot there. And you don't feel much pain there. [South Asian female]

...because winter is the time when it really gets more and more kind of affected. [White British male]

Across all ethnic groups, the negative effects of pain on mental health were consistently mentioned and the participants expressed their mental state as brain fog, confused, stressed, dementia-like, trauma, and bad mood.

...it's just that as it [pain] progresses it was affecting my memory as well. [Black African female]

Similarly, participants across all ethnic groups mentioned how their pain negatively affected their relationship with family members.

**Theme 2: Approaches and Attitudes to Self-treatment and Management**

Participants across all ethnic groups expressed their dissatisfaction with the treatment they were currently receiving. They also described how they relied on self-management practices and on pacing themselves to manage their painful condition better. Thinking about the diagnosis of their painful conditions, some participants said pain was an unexpected diagnosis for them, while others expressed frustration about delays in having their condition diagnosed as such.

...I had to run to my GP on many occasions...to explain that I'm suffering with this pain and I want to get to the bottom of what it is...and the doctor said to me oh, you're still young. You're still in your 20s. You can't have this [painful condition]. [South Asian female]

Participants also expressed concerns about treatment effectiveness and how they were given different treatments and but remained unable to manage their pain effectively. Participants discussed how they developed the practice of self-medication.

...I now self-medicate myself according to the level of pain that I've actually got. [White British male]

For managing pain, a participant described medication practice with a cultural viewpoint.

...We tend to tolerate it perhaps in a different way, and adjust really the cultural issue of not using medications or tablets as, almost like sweets. So we tend to only use medication where it's absolutely necessary. [Black African male]

**Theme 3: Frustration and Embarrassment in Communicating About Pain With Others**

Communicating about pain with friends, family members, and health care professionals was described as challenging across all ethnic groups. One of the participants described how communicating pain history during consultations was particularly difficult.

...And having to do some consultation, I get irritated because asking me to check my joints...how would I know what to do? How do I...I can’t tell my progress in a week, in a month. I really can't unless I keep a diary of what's going on.... [Black African female]

South Asian participants, particularly women, shared feeling embarrassed when talking about their pain.

...I think it's the way you're brought up...some people find it embarrassing, that shouldn't be discussed with the rest of the family. [South Asian female]

Male participants also described how the image of masculinity in their culture and the need for preserving their self-image hindered them to talk about their pain, which led them to developing a negative reporting behavior.

...men are more resistant to expressing their medical conditions because they are so much...stronger and it's not supposed to be...like a man to complain about anything. [Southeast Asian male]

A White British participant shared a similar perspective on self-image but less directly linked to his cultural background compared to South Asian participants.

...But for me, talking to others is about managing my own self-image. Because...people see me in a particular way. And the fact that I'm unable to do certain things....reduces me in some way, in my own mind, to some extent. And so I tend not to talk. [White British male]

Black African participants mentioned pain was perceived as a disability in their culture, thereby reinforcing their negative reporting behavior.

...Disability is not something that is seen as something to talk about in our culture. You just want to hide things and just behave as if everything is okay. [Black African female]

With hiding disability being the norm, it also limited them to optimally manage their pain.
So that’s another problem. You’re not sure if using the aid will make you better. But then you don’t want to use it because of the attention it creates as well. So all this contributes to the mental struggle. [Black African female]

Communicating pain to health care professionals was found to be equally challenging.

...Just in general, I find it very hard to communicate with medical professionals where the pain is, what it feels like, the very fact that it’s even real. [White British female]

A Black African participant expressed how it could be more beneficial to speak to a health care professional with a similar cultural background.

...But like others are saying, honestly, if [my doctor] came from the same background as mine...he was African...I think it would have been better to explain how the pain was going. Because we’ve got the actual words to actually explain how the pain is like. [Black African female]

With regard to describing pain in culturally appropriate and understandable language or terms, the following 2 contrasting opinions were noted.

...I was lucky. My GP is of Asian background but he lived in Africa. So it was more like a fatherly conversation kind of thing. So that helped as well because he understood where I was coming from. [Black African female]

...I went to my GP a while back, the GP was Gujarati [South Asian] and I just didn’t feel comfortable disclosing my issues to him. [South Asian male]

Theme 4: Lack of Experience With Formal Pain Assessment Tools

Few participants had experience of completing pain self-assessment tools as part of their care, and those who had completed were unhappy because they thought pain reporting methods did not capture their pain situation comprehensively.

...there’s a picture of a person and you have to put a cross on the places where you've got pain. But...that just tells them there’s pain in that area. It doesn’t give them a good indication of how much pain, whether it’s worse in certain areas than others. [South Asian female]

...So being told to grade the pain to a physician is very, very difficult for me to do. [Black African female]

...The GP was instantly like the others, just saying, is it every day, on a scale of 1 to 10 what is it? And you just feel so rushed that you don’t get a chance to explain that no, it’s not every day but some days it’s bad at a certain time in the day. [White British female]

Some participants who had experience of reporting pain using the Manchester Digital Pain Manikin app in our feasibility study (Ali SM et al, unpublished data, January 2023) described their experience as follows:

...I felt like describing my pain to someone. I thought someone’s listening to me, someone’s understanding it. [South Asian female]

...something like this [a smartphone app] would be very ideal in the context that it would be very confidential. I would have the opportunity to input area of my pain to get better advice. [South Asian male]

User Requirements for Digital Pain Self-report Tools

In total, 21 user requirements across 4 categories emerged from the synthesis of participants’ views on what pain aspects and app features were important (see Table 3). Nine requirements were consistent across ethnic groups, while 12 were only mentioned during one of the ethnicity-specific workshops. Below, we summarize per category similarities and differences in requirements between ethnic groups and how differences were discussed during the member-checking workshop; we did not discuss similarities during this workshop.
Table 3. User requirements for digital pain self-report tools.

<table>
<thead>
<tr>
<th>Pain aspects/app features</th>
<th>Ethnicity-specific workshops&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Member-checking workshop&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Remarks&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>South Asian</td>
<td>Black African</td>
<td>White British</td>
</tr>
<tr>
<td><strong>Location-specific pain aspects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pain quality (eg, stabbing, throbbing)</td>
<td>Yes&lt;sup&gt;d&lt;/sup&gt;</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Location-specific pain intensity</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Pain radiation</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Pain layers or depth&lt;sup&gt;f&lt;/sup&gt;</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>New pain</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Pain timing/duration</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Non-location-specific pain aspects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pain causes and aggravating factors (ie, factors that cause or increase)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Pain impact (ie, interference with other activities or consequences)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Pain management strategies</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Semistructured diary field (with headings as suggestions for what to record in this field)&lt;sup&gt;f&lt;/sup&gt;</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Reasons for not reporting pain</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td><strong>App features: Feedback and output</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feedback of previous pain reports&lt;sup&gt;f&lt;/sup&gt;</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Pain management guidance based on pain reports</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>App features: Look and feel</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Available for any digital device</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Flexible reporting frequency</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Multiple languages</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Intuitive color scheme linked to pain intensity scores</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Manikin zoom-in function (by finger pinch)</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Manikin body sides (minimum front, back, lateral sides)</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Manikin detail</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>
**Summary of Findings**

We conducted 3 web-based focus groups followed by a user requirement exercise with people from different ethnic backgrounds living with a chronic pain condition. We found many similarities in how the participants described their experience of living with pain; how pain management is still suboptimal; and how it is challenging to communicate about pain with their friends, family members, and health care professionals. People from non-White ethnic backgrounds had different beliefs and perceptions on pain compared to those from White backgrounds, which resulted in internalizing stigma and developing a negative attitude toward medication and pain reporting. Despite these differences, participants across ethnic backgrounds agreed on which aspects of pain reporting were important to self-report, such as pain quality, pain causes, activities that seemed to aggravate their pain to which South Asians also agreed during the member checking.

**Location-Specific Pain Aspects**

During 2 ethnicity-specific workshops (South Asian and Black African) and the member-checking workshop, participants identified pain layers (eg, skin, muscle, bone) as an important aspect. However, when we showed mock-ups of how this could be implemented in the Manchester Digital Pain Manikin, participants reported that this might overcomplicate pain reporting, suggesting that translating this requirement into a functionality may not be straightforward. When discussing pain radiation and pain duration as aspects during the member-checking workshop, participants agreed these were relevant for the digital pain self-report tool.

**Non–Location-Specific Pain Aspects**

All participants mentioned that they would be motivated to regularly self-report their pain if this would enable them to manage their pain better. Across workshops, participants described reporting of pain causes or aggravating factors crucial in this context. However, we found during the focus groups that the type of perceived pain causes varied across groups. Only White British participants suggested a semistructured diary field to capture information about diet, mood, physical activity, and level of medication, which participants from the other 2 ethnic groups appreciated during the member-checking workshop when shown mock-up screens for this functionality. They additionally suggested that such a diary field could be linked to a specific pain location to enable location-reporting of factors associated with pain (eg, perceived pain causes).

**Feedback and Output**

Participants wanted summaries of their pain reports, which in their view would enable them to track pain fluctuations in relation to changes in management and coping strategies. Participants confirmed this requirement during the member-checking workshop after seeing mock-ups of the pain summary reports while also sharing additional thoughts on how best to summarize the changes in pain, medication use, and coping strategies. Black African and White British participants also suggested that personalized data-informed messages could, for example, encourage people to refrain from undertaking

---

### Table: Pain aspects/app features

<table>
<thead>
<tr>
<th>Pain aspects/app features</th>
<th>Ethnicity-specific workshops&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Member-checking workshop&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Remarks&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manikin personalization (eg, gender-specific)&lt;sup&gt;f&lt;/sup&gt;</td>
<td>South Asian: No, Black African: Yes, White British: Yes</td>
<td>Agreed</td>
<td>Increased relevance to user; more life-like</td>
</tr>
</tbody>
</table>

<sup>a</sup>Requirements that were mentioned during the ethnicity-specific workshop are represented as yes and those that were not mentioned as no.  
<sup>b</sup>Consistently reported requirements across all ethnicity-specific workshops were not discussed during the member-checking workshop and are therefore shown as not discussed. For requirements that were discussed, agreement across participants is represented as agreed and lack of clear agreement as not agreed.  
<sup>c</sup>Summary of the illustrative participant comments noted during breakout groups (in ethnicity-specific workshops) to clarify why people considered certain pain aspects and app features important.  
<sup>d</sup>Yes means a pain aspect or app feature was mentioned during a particular ethnicity-specific workshop.  
<sup>e</sup>Agreed means participants agreed on its importance during the member-checking workshop.  
<sup>f</sup>Requirement presented as a mock-up screen for the Manchester Digital Pain Manikin app to gather further thoughts on how the requirement could be translated into a functionality.
feedback of previous pain reports, and availability of a digital device for pain management. However, we found differences in requirements related to language (eg, translated in-app instructions, culturally appropriate pain terminologies) and that people did not always agree on how best to translate requirements into reporting functionality (eg, pain layers/depth). Addressing these differences when developing digital pain self-report tools will enhance their cross-cultural acceptability and contribute to more equitable pain management and outcomes by reducing pain reporting barriers across ethnic groups.

Relation to Other Studies

We found that gender stereotypes and associated stigma, which may vary across cultures, influenced people’s pain experience and reporting behavior negatively. For example, Black African female participants in our study said that pain is viewed as a disability, leading to negative disclosure behavior (ie, people are less likely to report their pain). This aligns with findings from a review by Bakhshaie et al [9] in 2022 who suggested that stigma internalization (eg, when somebody links their disability to their personality) in Black individuals results from the interplay between interpersonal, community, and societal factors, which in turn is related to discrimination and societal injustice [9]. Similarly, South Asians indicated that pain among women is considered inevitable because of women’s household responsibilities. Owing to the conventional gender roles, men may be less willing to report pain and more willing to endure it [29]. This finding is in line with those reported in other studies [30,31] that specific expectations evoked by gender, ethnicity, nationality, or religion may further complicate pain experience.

We found that there was a general criticism among participants about single-rating scales and other existing tools. One issue they highlighted was that they found those tools too simplistic for their complicated pain situation. The identification of different pain aspects, for example, intensity, quality, frequency, duration, and their temporal aspects; pain causes; and impacts are consistent with recommended core outcome measures for chronic pain [32]. In addition, assessment tools for pain self-reporting may affect the patient-provider encounter and contribute to more equitable pain management and outcomes by reducing pain reporting barriers across ethnic groups.

Limitations of This Study

One limitation of our study was that the samples for each of the ethnicity-specific workshops were relatively small and may not have reflected the wide range of cultural diversity within a specific ethnic group. For example, the Pakistani culture comprises numerous ethnic groups such as Punjabis, Kashmiris, Sindhis, and Muhajirs. Therefore, specific pain belief and pain self-reporting needs within ethnic subgroups and examining to what extent these beliefs and needs are common in such subgroups across countries (eg, Punjabis living in Pakistan, India, and the United Kingdom) is an area of future research.

Another limitation was that only people who spoke and understood English and who had access to a digital device and the internet could take part in the workshops. This may have further reduced the diversity of our sample. For example, people with a disability or those who are older are less likely to use the internet [36]. Similarly, although South Asians are more likely to experience chronic pain [5], not all may be sufficiently proficient in English to participate in group discussions, thereby limiting the generalizability of our findings. Future studies could therefore consider conducting in-person interviews or focus group discussions in people’s own language (eg, Urdu) at a convenient place (eg, a community center). In addition, as these people are more likely to represent less affluent groups, engaging with them would help us examine the intersectional considerations (related to ethnicity; eg, income level, occupation type, education level) within a specific ethnic group.

Implications for Developing Cross-culturally Acceptable Digital Pain Self-report Tools

People across ethnic groups mostly agreed on what were relevant and important aspects of pain, which included pain causes. However, differences in perceived pain causes between ethnicities, such as food, weather, and gender norms, should be acknowledged to facilitate culturally relevant pain self-reporting that supports people with self-managing their pain. Similarly, digital pain self-report tools such as smartphone-based pain manikins showed potential in overcoming challenges of communicating pain with health care professionals, especially for people from non-White ethnic backgrounds, which suggests that pain drawings may have clinical utility [37]. However, this requires cultural (eg, culturally appropriate pain terminologies) and linguistic (eg, translated instructions in users’ native language or use of audio/video instead of text) compatibility across a wide range of ethnic backgrounds.

In addition to these features, digital tools incorporating a pain manikin should offer the option of personalizing the body shape [17]. However, our experience from the member-checking workshop showed that it is not straightforward to translate user requirements related to manikin personalization into app functionalities that meet people’s expectations. Further, adding more functionalities to increase cultural and gender appropriateness needs balancing against increasing the complexity of the pain self-report tool as intended to avoid creating barriers for other potentially disadvantaged groups (eg, those with lower digital literacy levels or limited manual dexterity). Lastly, offering personalization options may affect the measurement properties of digital manikins and how we interpret manikin drawings and the data derived from them. Developers of digital manikins and researchers should further explore how best to address the need for manikin personalization and its impact on data collection, analysis, and interpretation.

We need innovative user-centered prioritization techniques to facilitate the development of equitable digital pain and other health assessment tools. Currently, methods for prioritizing requirements, which emerged from an increased need to involve
stakeholders in developing software and information systems [38], are commonly based on majority votes, for example, the Top10, cumulative voting, and numerical assignment [39]. However, in our study, we found some user requirements that were only relevant to a specific minority group, and existing prioritization techniques insufficiently encourage developers to appreciate these.

**Conclusion**
Exploring the views of people from different ethnic backgrounds generated new insights into their pain experiences and challenges in communicating their pain. There were cultural differences in perceived causes of pain, self-management strategies, and their reporting behavior because of gender norms and the stigma associated with pain. Moreover, there were differences in language requirements. Acknowledging and addressing these differences is important for the development of cross-culturally acceptable digital pain self-report tools, which in turn will contribute to reducing inequities in pain treatment and outcomes.

**Acknowledgments**
The University of Manchester Medical Research Council’s Confidence in Concepts scheme 8 funded this study.

**Conflicts of Interest**
BJ and SMA are employees of uMotif, and BJ is a cofounder. uMotif Limited is the technology partner that supported the design and development of the Manchester Digital Pain Manikin app. WGD has received consultancy fees from AbbVie and Google, which are unrelated to this work.

Multimedia Appendix 1
Study flyer.
[PDF File (Adobe PDF File), 256 KB - humanfactors_v10i1e42177_app1.pdf ]

Multimedia Appendix 2
Topic guide.
[DOCX File, 39 KB - humanfactors_v10i1e42177_app2.docx ]

Multimedia Appendix 3
Additional illustrative quotes.
[DOCX File, 15 KB - humanfactors_v10i1e42177_app3.docx ]

**References**


Text Messages Exchanged Between Individuals With Opioid Use Disorder and Their mHealth e-Coaches: Content Analysis Study

Yerina S Ranjit1, PhD; Warren M Davis1, PhD; Andrea Fentem2, MA; Raven Riordan2, MA; Rikki Roscoe1, MA; Patricia Cavazos-Rehg2, PhD

1Department of Communication, University of Missouri, Columbia, MO, United States
2Department of Psychiatry, Washington University School of Medicine in St Louis, St Louis, MO, United States

Corresponding Author:
Yerina S Ranjit, PhD
Department of Communication
University of Missouri
207 Switzler Hall
Columbia, MO, 65211
United States
Phone: 1 573 882 4431
Email: ranjity@missouri.edu

Abstract

Background: Opioid use disorder (OUD) has affected 2.2 million people in the United States. About 7.2 million people reported using illicit drugs in 2019, which contributed to over 70,000 overdose deaths. SMS text messaging interventions have been shown to be effective in OUD recovery. However, the interpersonal communication between individuals in OUD treatment and a support team on digital platforms has not been well examined.

Objective: This study aims to understand the communication between participants undergoing OUD recovery and their e-coaches by examining the SMS text messages exchanged from the lens of social support and the issues related to OUD treatment.

Methods: A content analysis of messages exchanged between individuals recovering from OUD and members of a support team was conducted. Participants were enrolled in a mobile health intervention titled “uMAT-R,” a primary feature of which is the ability for patients to instantly connect with a recovery support staff or an “e-coach” via in-app messaging. Our team analyzed dyadic text-based messages of over 12 months. In total, 70 participants’ messages and 1196 unique messages were analyzed using a social support framework and OUD recovery topics.

Results: Out of 70 participants, 44 (63%) were between the ages of 31 and 50 years, 47 (67%) were female, 41 (59%) were Caucasian, and 42 (60%) reported living in unstable housing conditions. An average of 17 (SD 16.05) messages were exchanged between each participant and their e-coach. Out of 1196 messages, 64% (n=766) messages were sent by e-coaches and 36% (n=430) by participants. Messages of emotional support occurred the most, with 196 occurrences (n=8, 0.8%) and e-coaches (n=187, 15.6%). Messages of material support had 110 occurrences (participants: n=8, 0.7%; e-coaches: n=102, 8.5%). With OUD recovery topics, opioid use risk factors appeared in most (n=72) occurrences (patient: n=66, 5.5%; e-coach: n=6, 0.5%), followed by a message of avoidance of drug use 3.9% (n=47), which occurred mainly from participants. Depression was correlated with messages of social support (r=0.27; P=.02).

Conclusions: Individuals with OUD who had mobile health needs tended to engage in instant messaging with the recovery support staff. Participants who are engaged in messaging often engage in conversations around risk factors and avoidance of drug use. Instant messaging services can be instrumental in providing the social and educational support needs of individuals recovering from OUD.

JMIR Hum Factors 2023;10:e37351 doi:10.2196/37351

Keywords

opioid use disorder; opioid; opium; overdose; drug; substance use; content analysis; text message intervention; text message; text messaging; mobile health; mHealth; social support; e-coach; counseling; mental health; depression; recovery support; eHealth; digital health
**Introduction**

**Background**

Opioid use disorder (OUD) has affected 2.2 million people in the United States. About 7.2 million people report using illicit drugs such as heroin, fentanyl, and prescription opioids, which has contributed to over 70,000 overdose deaths in 2019 [1,2]. This formidable crisis has multiple health and social implications for people with OUD, who are at a high risk of comorbidities, including HIV and mental health disorders, and have higher rates of mortality compared to the general population [3,4]. OUD is also associated with adverse social outcomes including being incarcerated, homelessness, and experiencing social stigma [5-7].

OUD is treated with medication for opioid use disorder (MOUD) using opioid agonist therapy with methadone or buprenorphine [8,9]. MOUD is generally administered over a period of time that involves medically supervised withdrawal, maintenance, and continued psychosocial support for patients [9]. The treatment period differs for each patient depending on the severity of dependence and other medical factors [9]. Behavioral therapy and counseling designed to prevent relapse and support patients are considered the standard of care in addition to pharmacological intervention [9]. Because OUD is a chronic condition that requires pharmacological intervention with ongoing behavioral intervention during recovery, communicating with patients regularly is important to sustain improved health outcomes [9].

Mobile health (mHealth) intervention or the use of mobile phones to improve health holds immense promise in the treatment of chronic conditions including OUD [9-12]. Evidence suggests that mobile phone interventions, including SMS text messaging reminders, have a positive impact on self-management of chronic illnesses [13]. Research shows that patients with OUD spend similar amounts of time on the internet as the general population, and there is high acceptability of mHealth interventions among this population [12,14,15]. A reason that mobile phones are effective in managing chronic illnesses is that users experience a level of social support when information is provided at regular intervals via SMS text messaging [16]. Effectiveness of SMS text messaging programs in initiating treatment as well as preventing relapse has been demonstrated in recent studies [15,17]. Additionally, 70% of the patients at a primary care office-based buprenorphine treatment expressed interest in receiving supportive SMS text messaging, in addition to messages pertaining to the risk of relapse [18]. This suggests there is interest among persons with OUD in receiving social support via mHealth. However, limited studies have examined the content of interpersonal communication between patients in OUD treatment and their providers that have occurred via SMS text messaging.

**Role of Social Support in Patient-Provider Communication in OUD Treatment**

Social support occurs when messages, verbal or nonverbal, express directly or indirectly that someone is valued and cared for [19]. Jacobson [20] categorized social support into emotional, cognitive, and material support. Emotional support refers to “behavior that fosters feelings of comfort and leads an individual to believe that he or she is admired, respected, and loved, and that others are available to provide caring and security” [20]. Cognitive support may include information, knowledge, or advice that can help an individual understand their world and adjust to changes, and material support is defined as goods and services that can help solve practical problems [20]. Hence, messages of social support often convey information, emotion, or referral to help someone to manage and reduce uncertainty [16]. Individuals in OUD treatment and recovery experience many uncertainties about the treatment process and outcomes, their personal lives, and social reintegration [21,22]. Additionally, chronic relapse is common in OUD, and research show that social support is crucial for relapse prevention and abstinence [17,23-25]. For example, a study by Polenick et al [26] showed that women undergoing OUD treatment who measured high on loneliness were more likely to start using illicit drugs during recovery compared to those who had greater social support. Additionally, informational support and feeling of closeness played a significant role in the recovery process for pregnant women with OUD and decreased their substance use [27,28]. Hence, there is ample evidence that social support can reinforce the benefits of medication treatments for OUD [29].

Additionally, sustained communication between patients and providers is important for effective treatment outcomes due to the chronic nature of OUD, especially because people with OUD may experience loss of meaningful relationships due to addiction [9,22,24,30]. Supportive relationships between persons with OUD and providers can be defined by mutual trust, respect, and understanding void of prejudice, negative attitudes, or discrimination [9,31]. With the advancement of digital technology, many health interventions effectively use the instant messaging feature on mobile phones to convey messages of social support [32]. The advantages of using SMS text messaging include scalability, relative low cost, and the ability to tailor and personalize messages [32]. While the impact of social support on other health outcomes have been widely examined, social support via mHealth intervention for OUD treatment and recovery has not been well researched. Hence, the purpose of this study is to explore the content of SMS text messaging between participants in a recovery program, with a focus on social support, treatment-related messages, and their relationship with other health outcomes in this population. Understanding social support themes in SMS text messaging exchanges with providers will help us design future text-based mHealth interventions specific to the needs of OUD patients.

**This Study**

This study is part of a larger mHealth intervention titled “uMAT-R,” a supplemental support tool to improve adherence to OUD treatment and recovery. The parent study used a mobile app to provide educational content and psychological support for people in OUD recovery programs in the Greater St. Louis Area. uMAT-R also contains modules to support recovery efforts including medication and appointment reminders, and
community resources. The primary feature of uMAT-R is the ability for patients to instantly connect with a recovery support staff member or an “e-coach” via in-app messaging. There were 4 e-coaches in this study, who hold bachelor- and master-level degrees in backgrounds including clinical psychology, public health, and social work. e-Coaches received introductory training in person-centered coaching techniques, motivational interviews, patient-centered therapy techniques, and crisis intervention [33,34]. According to the study protocol, after enrollment, the assigned e-coach sent an initial scripted message to let the participants know that they were available for support at any time. If the participants did not respond to the initial message, the e-coach provided a check-in message once a week to let them know that an e-coach was still available. After the initial message, if participants replied to the e-coaches, the response messages from e-coaches were unscripted and tailored to address the specific needs of individual participants. During working hours, e-coaches were tasked with responding immediately after receiving messages from participants. e-Coaches used the first names of the participants while responding to the messages in order to personalize the messages. Each week, e-coaches meet as a team to discuss caseloads and the progress of their clients. They were encouraged to reach out to the project manager or principal investigator whenever needed. If a crisis message was received outside the e-coach’s working hours, the project manager reviewed the message and alerted the e-coach if the message needed to be attended to right away. If the e-coach is unavailable, the project manager or the principal investigator directly responded to the client via messaging or follow-up via phone if there is a concern regarding safety.

This study analyzed these SMS text messages exchanged between participants undergoing OUD recovery and their e-coaches. Additionally, the psycho-educational content within uMAT-R covers topics such as avoiding drug use, dependency, triggers, risk factors, and tips on developing and maintaining healthy alternative habits.

In order to understand the content of these dyadic or two-way messages, this study explores the following research questions: (1) What proportion of in-app social support messages, including messages about emotional, informational, and material support, were exchanged by participants and their e-coaches? (2) What proportion of messages related to the recovery process covered in uMAT-R modules, including messages about avoiding drug use, relapse, triggers, healthy habits, dependency, and risk factors, were exchanged by participants and their e-coaches? (3) How is the mental health condition of OUD recovery participants related to the in-app messaging? (4) What is the nature of dyadic message exchange between participants and e-coaches around topics of social support and OUD recovery?

### Methods

#### Sample

For the parent study, participants were recruited from various types of facilities such as OUD outpatient and inpatient programs, recovery homes, hospital settings, medication for addiction treatment clinics, and a clinic that primarily supports pregnant and postpartum women in recovery. Participants were either attending treatment voluntarily or court mandated. Participants were eligible to partake in the uMAT-R mobile app study if they met the following criteria: (1) if they had ever received a formal OUD diagnosis, (2) were currently receiving opioid addiction recovery treatment at one of the above settings, (3) were 18 years or older, (4) were a US resident, (5) were fluent in English, and (6) owned a smartphone with an iOS or Android operating system.

The in-app message exchange between the e-coaches and the participants using the mobile app was retrieved intermittently with the participants’ permission. We retrieved messages over a 12-month period, from 2019 to 2020.

In total, 80 unique dyadic sets of communication occurring between e-coaches and individuals in recovery were identified in the initial data retrieval process, with 1666 individually sent messages present within the dyadic texts. After reviewing the data, the research team removed messages that were unidirectional, such as messages sent by e-coaches without a response from the participants. As a result, the final sample totaled 70 unique dyadic sets, with 1196 individually sent messages. Because the purpose of the study was to examine the impact of the interactive messages, nondyadic messages were excluded from the study. Each dyadic set had a varying number of messages exchanged between OUD patients and e-coaches. The research team aimed to identify the prevalence of thematic elements of social support and OUD-related topics within the data.

#### Ethics Approval

This study was approved by the Washington University in St. Louis’ Institutional Review Board (#210805132 and #201910161).

#### Statistical Analysis

The prevalence of themes pertaining to social support and topics related to OUD recovery were explored using summative content analysis. In this approach, the research process incorporates identifying and quantifying certain words or thematic elements in a text to understand the context in which they are being used [35]. A summative approach to qualitative content analysis differs from quantitative content analysis in that it goes beyond simply counting words. That is to say, this summative approach includes the process of interpreting content, often referred to as latent content analysis [35]. Additionally, to understand the relationship between mental health outcomes and the nature of the messages exchanged, we conducted a Pearson correlational analysis between individual participant’s scores on depression and anxiety and the type of message (social support and OUD related) sent by each individual.

#### Codes

With the primary goal of determining the extent to which messages of social support and OUD-related topics were present in the SMS text messaging interactions, the research team developed a codebook to examine the messages between participants in OUD recovery and their e-coaches. Each case, or individual message, was coded for the following: emotional support messages or messages that foster feelings of comfort...
and leads the individual to believe that he or she is admired, respected, and loved and that others are available to provide caring and security [36]; cognitive support messages, or informational support, is knowledge or advice that helps the individual to understand his or her world and to adjust to changes within it [36]; and material support messages or messages about goods and services that help solve practical problems [20,36].

The larger intervention aimed to improve the knowledge of individuals who are in recovery around 7 treatment domains. Codes were, hence, created to identify these OUD treatment and recovery domains that included messages of (1) dependency or high level of tolerance for opioid use and a mention of withdrawal symptoms such as diarrhea, sleeplessness, restlessness, irritability, and psychomotor agitation [37]; (2) craving or the desire for more opioid use, as well as the desire to avoid the withdrawal [38]; (3) relapse or recurrence of drug use after a period of abstinence [39]; (4) risk factors or anything that may contribute to use of opioids including alcohol use, psychiatric conditions, home, family, and social environment that will encourage opioid or drug use [39]; (5) triggers or environmental factors, including people or places or moods that trigger drug use [39]; (6) avoidance or self-control and motivation, commitment, and willingness to stay away from drugs, even in difficult situations [40]; and (7) healthy alternative habits or alternatives used in recovery to minimize relapse such as exercising, mindfulness meditation, positive reframing, journaling, and so on [41]. Finally, the codes included mobile app usability issues or problems participants faced that were related to the mobile app.

Coding Procedure

Each individually sent in-app message served as the primary unit of analysis. This approach allowed us to identify the variable presence on the individual and dyadic level messages. Two coders analyzed the content for either the presence or absence of variables. The training of coders occurred using messages not included in the final coding sample. Both coders engaged in a reliability training over several weeks. In instances of disagreements regarding the interpretation of the thematic content, the research team discussed the concepts in the codebook and compared the data until all discrepancies were resolved. Approximately 173 individual messages were used in reliability training, roughly equating to 11% of the final coding sample, with all variables coded in a binary manner (ie, presence or absence). In coding the final messages sample of 70 dyadic text exchange sets, the research team continued with this process of consensus coding.

Results

The results showed that 67% (n=47) of the participants in the study identified as female and 33% (n=23) as male. In regards to race and ethnicity, 59% (n=41) were Caucasian, 20% (n=14) African American, and 1% (n=1) Hispanic or Latinx. More than half of the participants reported living in unstable housing conditions and 57% (n=38) were unemployed. See Table 1 for detailed demographic information.

With regard to in-app text messages, out of 70 unique dyadic sets, a minimum of three and a maximum of 87 messages were exchanged between e-coaches and participants. Messages were initiated by e-coaches in most conversations (n=63, 90%), while participants initiated 10% (n=7) of messages. Overall, an average of approximately 17 (SD 16.05) messages per conversation were exchanged. Out of 1196 messages, 64% (n=766) of the messages were sent by the e-coaches and 36% (n=430) were sent by participants. e-Coaches sent an average of 4.1 (SD 5.9) more messages than the participants. Overall, the participants sent an average of 6.6 (SD 7.8) messages per conversation, and e-coaches sent an average of 10.7 (SD 9.4) messages per conversation. The maximum number of messages sent by participants in 1 conversation was 39 and the minimum was 1, and the maximum number of messages sent by e-coaches was 50 and the minimum was 2 per conversation.

In reference to social support messages, the results showed that emotional support surfaced the most in these dyadic conversations, with 196 occurrences between participants (n=9, 0.8%) and e-coaches (n=187, 15.6%). Material support also occurred at a relatively high frequency, with 110 occurrences between participants (n=8, 0.7%) and e-coaches (n=102, 8.5%). In regard to OUD treatment topics, opioid use risk factors appeared the most with 72 total occurrences, which were mostly sent by participants (n=66, 5.5%) than by e-coaches (n=6, 0.5%), followed by messages of avoidance with 47 total occurrences all sent by patients. Additionally, messages about following a healthy lifestyle were also mostly sent by participants (n=22, 1.8%), with a total of 24 occurrences. Details of message distribution and textual examples are provided in Figure 1 and Table 2.

Additionally, we conducted a Pearson correlation to examine the relationships of the demographic variables, mental health conditions of participants, and adherence to the recovery program to the types of messages presented in Table 3. We found that participants who measured high on depression also sent a higher number of messages of all types of social support (r=0.27; P=.02). Examining the number of messages only from participants, messages of OUD topics were correlated with messages of social support (r=0.42; P<.001), and messages of app usability (r=0.75; P<.001).

Examining the dyadic exchange between participants and e-coaches, we found that messages of social support from e-coaches were highly correlated with messages from participants, especially around messages of social support (r=0.34; P<.001) and messages around OUD topics (r=0.75; P<.001).

In Textbox 1, we provide some examples of qualitative data on the communication pattern that support the correlation between individuals with OUD seeking help and e-coach providing support.
Table 1. Demographic variables of patients in U-MAT-R program (N=70).

<table>
<thead>
<tr>
<th>Participant characteristics</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
</tr>
<tr>
<td>19-20</td>
<td>19 (27)</td>
</tr>
<tr>
<td>31-50</td>
<td>44 (63)</td>
</tr>
<tr>
<td>51 and older</td>
<td>3 (4)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>23 (33)</td>
</tr>
<tr>
<td>Female</td>
<td>47 (67)</td>
</tr>
<tr>
<td><strong>Race and ethnicity</strong></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>41 (59)</td>
</tr>
<tr>
<td>African American</td>
<td>14 (20)</td>
</tr>
<tr>
<td>Hispanic/Latinx</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Other</td>
<td>1 (1)</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
</tr>
<tr>
<td>Completed high school</td>
<td>58 (83)</td>
</tr>
<tr>
<td>College and above</td>
<td>9 (13)</td>
</tr>
<tr>
<td><strong>Employment</strong></td>
<td></td>
</tr>
<tr>
<td>Full-time</td>
<td>16 (23)</td>
</tr>
<tr>
<td>Part-time</td>
<td>13 (19)</td>
</tr>
<tr>
<td>Not employed</td>
<td>38 (57)</td>
</tr>
<tr>
<td><strong>Housing situation</strong></td>
<td></td>
</tr>
<tr>
<td>Unstable</td>
<td>42 (60)</td>
</tr>
<tr>
<td>Stable</td>
<td>25 (36)</td>
</tr>
<tr>
<td><strong>Number of SMS text messaging between participants and e-coach, mean (SD)</strong></td>
<td>17 (16.05)</td>
</tr>
</tbody>
</table>

Figure 1. Distribution of 1196 text messages exchanged between patients and providers in the opioid use disorder (OUD) recovery program.
Table 2. Distribution of messages by content category.

<table>
<thead>
<tr>
<th>Types of messages</th>
<th>Message count (N=1196), n (%)</th>
<th>Sample text</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Participants</td>
<td>e-Coaches</td>
</tr>
</tbody>
</table>
| Emotional support | 9 (0.8)       | 187 (15.6) | • Provider: “Hello! I just wanted to check in with you during these times of self-isolation and uncertainty. How have you been doing? Is there anything I can do to help?”
|                   | 187 (15.6)   | 9 (0.8)    | • Patient: “Hello. Trying to hang in there. It's hard being quarantined, as I'm sure you know. Thanks for asking! I hope you are doing well!” |
| Cognitive support | 2 (0.1)       | 55 (3.6)   | • Provider: “Here's a good website for some additional tips on how to manage tantrums as well: https://kidshealth.org/en/parents/tantrums.html?”
|                   | 55 (3.6)     | 2 (0.1)    | • Patient: “Do you have any medical resources in regards to getting on PrEP?” |
| Material support  | 8 (0.6)       | 102 (8.5)  | • Provider: “But let me know if you have any questions! You are still able to finish your baseline survey that was sent to your email address and once that's done we can send you your $30 Walmart gift card!”
|                   | 102 (8.5)    | 8 (0.6)    | • Patient: “thank you for getting back to me. I looked into both of those resources already prior to reaching out. I was looking for additional options.” |
| Cravings          | 2 (0.1)       | 0 (0)      | • Patient: “I am having really bad Xanax cravings. One of the reasons is my anxiety is really high today.” |
| Risk factors      | 66 (5.5)      | 6 (0.5)    | • Provider: “Hi there! Just wanted to check in and see how you're doing, I know you've had a very stressful and long week. If you ever want to talk, just know I'm here!”
|                   | 6 (0.5)       | 66 (5.5)   | • Patient: “MY choices to not let go... how do u let go of someone u kno isnt good for your entire life... but u dnt wanna leave em i dnt kno if it's outta fear or loneliness ... this lifestyle i can't make it” |
| Trigger           | 6 (0.5)       | 1 (0.1)    | • Provider: “I'm sure that was both emotionally and physically exhausting. I'm glad you're not at work today though and hope you're getting plenty of rest. Triggers like that can be difficult to deal with, no matter how far along you are in your recovery. Acknowledging and processing those triggers the way you did is great, I'm glad to hear you're doing well!”
|                   | 1 (0.1)       | 6 (0.5)    | • Patient: “hi *****, I'm ok. I'm having a lil issue this week, negative thoughts, not about using but I was in jail last two weeks. things had gone great for nearly two months. since I been home [though] that's changed. i feel I'm still doing great personally, but it seems like things out here went downhill while I was gone. at treatment and here at the house. [There’s] a lot of negativity now and distractions and it has me thinking of changing it up. moving and or switching treatments or at least switching to nights or whatever. i know I could use help or advice, so I thought I'd run it by you. i hope u r doing well.” |
| Avoidance         | 47 (3.9)      | 0 (0)      | • Patient: “I did my boundaries assignment and I read all the stuff...I got one [haven’t] been the best at setting boundaries but over the years I’ve gotten better.” |
| Healthy lifestyle | 22 (1.8)      | 2 (0.2)    | • Provider: “Today’s tip (found on your home screen) talks about a coping plan. Who or what is a part of your coping plan?”
|                   | 2 (0.2)       | 22 (1.8)   | • Patient: “I did my boundaries assignment and I read all the stuff...I got one haven’t been the best at setting boundaries but over the years I’ve gotten better” |
| App usability     | 19 (1.6)      | 91 (7.6)   | • Provider: “Hi! I wanted to let you know that my coworker [e-coach] will be taking over our messages for a while. Please feel free to reach out to her through the app if you need anything or have any questions moving forward!”
|                   | 91 (7.6)      | 19 (1.6)   | • Patient: “Just talked with [e-coach] and had to reset my password. now on I tether to get the help I can use. thank you”

https://humanfactors.jmir.org/2023/1/e37351

JMIR HUM Factors 2023 | vol. 10 | e37351 | p.531

(page number not for citation purposes)
Table 3. Correlations among the number and types of message exchanges, mental health, and demographic variables.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
<th>(10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r$</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P$ value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$^{a}$</td>
</tr>
<tr>
<td>(2) Education</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r$</td>
<td>0.18</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P$ value</td>
<td></td>
<td>0.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3) Depression</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r$</td>
<td>$0.28^{b}$</td>
<td>0.08</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P$ value</td>
<td></td>
<td>0.02</td>
<td>0.53</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4) Anxiety</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r$</td>
<td>$0.31^{b}$</td>
<td>0.05</td>
<td>0.66$^{c}$</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P$ value</td>
<td></td>
<td>0.01</td>
<td>0.69</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(5) Adherence to the recovery program</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r$</td>
<td>$0.19$</td>
<td>$-0.07$</td>
<td>$-0.01$</td>
<td>$0.02$</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P$ value</td>
<td></td>
<td>0.14</td>
<td>0.59</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(6) Messages of social support from participants</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r$</td>
<td>$0.21$</td>
<td>0.14</td>
<td>$0.27^{b}$</td>
<td>0.19</td>
<td>$-0.06$</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P$ value</td>
<td></td>
<td>0.09</td>
<td>0.28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(7) OUD$^{d}$-related messages from participants</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r$</td>
<td>$0.13$</td>
<td>$-0.10$</td>
<td>$-0.05$</td>
<td>$-0.05$</td>
<td>0.05</td>
<td>$0.42^{c}$</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P$ value</td>
<td></td>
<td>0.31</td>
<td>0.44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(8) Messages of app usability from participants</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r$</td>
<td>$0.03$</td>
<td>$-0.08$</td>
<td>0.11</td>
<td>0.12</td>
<td>$0.29^{b}$</td>
<td>0.00</td>
<td>0.06</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P$ value</td>
<td></td>
<td>0.79</td>
<td>0.51</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(9) Messages of social support from e-coaches</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r$</td>
<td>$0.27^{b}$</td>
<td>$-0.05$</td>
<td>0.13</td>
<td>0.10</td>
<td>0.14</td>
<td>$0.34^{c}$</td>
<td>0.75$^{c}$</td>
<td>$-0.02$</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>$P$ value</td>
<td></td>
<td>0.02</td>
<td>0.68</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(10) OUD-related messages from e-coaches</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r$</td>
<td>$0.31^{b}$</td>
<td>0.09</td>
<td>0.11</td>
<td>0.05</td>
<td>0.02</td>
<td>0.10</td>
<td>$0.28^{b}$</td>
<td>$-0.09$</td>
<td>$0.57^{c}$</td>
<td>1</td>
</tr>
<tr>
<td>$P$ value</td>
<td></td>
<td>0.01</td>
<td>0.47</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(11) Messages of app usability from providers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r$</td>
<td>$0.37^{c}$</td>
<td>0.15</td>
<td>0.09</td>
<td>0.09</td>
<td>0.14</td>
<td>0.19</td>
<td>0.17</td>
<td>0.22</td>
<td>$0.28^{b}$</td>
<td>$0.37^{c}$</td>
</tr>
<tr>
<td>$P$ value</td>
<td></td>
<td>0.002</td>
<td>0.24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^{a}$Not applicable.

$^{b}$Correlation is significant at the .05 level (2-tailed).

$^{c}$Correlation is significant at the .01 level (2-tailed).

$^{d}$OUD: opioid use disorder.
Principal Findings

This study conducted a content analysis of in-app text messages between individuals with OUD and their e-coaches in order to understand social- and OUD-related support needs of this population. Understanding the content of such text-based conversations can be instrumental in guiding the content of future text-based interventions.

The results of this study showed that in this intervention, e-coaches initiated conversation most of the time, but there were instances where the participants were the first ones to send a message. As a part of the design and the protocol of the larger mHealth intervention (uMAT-R), the support staff informed the participants that they would receive their first in-app message. As a part of the design and the protocol of the larger mHealth intervention (uMAT-R), the support staff informed the participants that they would receive their first in-app message. A possible reason for this could be the co-occurrence of multiple instances where the participants were the first ones to send a message. As a part of the design and the protocol of the larger mHealth intervention (uMAT-R), the support staff informed the participants that they would receive their first in-app message from an e-coach, after which they needed to respond to indicate the active functioning of their in-app message feature. The findings also showed that e-coaches sent more messages in each conversation compared to participants, showing that these messages were actively reaching out and responding to the messages sent by participants. The average number of messages exchanged between e-coaches and the participants showed a healthy trend, without overindulgence.

Further, the results showed that emotional support was the most occurring theme out of the 3 types of social support in the messages. Messages of emotional support from e-coaches occurred the most, followed by messages of material support from e-coaches. Higher number of messages seeking social support from individuals with OUD corresponded with a higher number of support messages sent by e-coaches. Collectively, these findings indicate that the e-coaches tasked with messaging participants reciprocated to the volume of messages sent by participants, a sign of effective interpersonal communication, important for the overall well-being of individuals [42]. While we recognize that studying reciprocity qualitatively was beyond the scope of this study, some qualitative examples of dyadic conversations show that e-coaches often acknowledged the contexts and experiences shared by participants in their response messages, providing evidence for “active listening,” a core value prescribed within the practice of motivational interviewing, for which the e-coaches had been trained prior to engaging in their supportive roles [33, 43].

Additionally, the results showed that e-coaches sent a higher number of messages of social support than did the participants. A possible reason for this could be the co-occurrence of multiple instances where the participants were the first ones to send a message. As a part of the design and the protocol of the larger mHealth intervention (uMAT-R), the support staff informed the participants that they would receive their first in-app message from an e-coach, after which they needed to respond to indicate the active functioning of their in-app message feature. The findings also showed that e-coaches sent more messages in each conversation compared to participants, showing that these messages were actively reaching out and responding to the messages sent by participants. The average number of messages exchanged between e-coaches and the participants showed a healthy trend, without overindulgence.

Furthermore, out of several OUD topics, participants sent messages with the theme of risk factors the most, followed by messages of avoiding drug use, possible triggers, and healthy alternative habits. The occurrences of these themes provide evidence for high informational needs due to uncertainties experienced by people in OUD recovery and the presence of various risk factors in their lives despite receiving MOUD, which collectively can contribute to relapse [44]. Hence, recovery and relapse-prevention programs should develop long-term interventions that focus on addressing these informational gaps as well as strategies to reduce perceived and actual risks.
Additionally, this study found that individuals with OUD with mental health needs, specifically those with depression, were more likely to engage in messages of social support. Previous meta-analysis shows that SMS text messaging interventions can have positive impact on managing depression [45]. In sum, because people with depression engage in seeking social support and because SMS text messaging interventions are effective in addressing mental health problems, future interventions should proactively incorporate themes that specifically address the mental health needs of people with OUD.

Overall, this study provides evidence that given an opportunity to engage in web-based two-way communication with their health care providers, people in such OUD recovery programs seek social support. These support needs, as shown in the study, are mostly that of emotional support followed by informational and material support. While measuring the outcome and the impact of the intervention was beyond the scope of this study, it clearly demonstrates that participants continually engage in support-seeking behavior. Programs that employ trained professionals and focus on addressing these needs through consistent communication, not only fill the support need gap but also create a sense of immediacy, defined as “perceived physical or psychological closeness,” which is an important factor in creating trust in providers [46]. Hence, while it is important to address the specific support needs through this web-based intervention, the very presence of a tool that allows for two-way and immediate connection could enhance participants’ perception of support. We recommend future studies to examine these perceptions and their impact further.

Limitations
While this study provides some important insights into digital intervention for OUD recovery, it is not without limitations. Because this was an exploratory study embedded in a larger parent study, we were not able to include attitudinal measures such as trust in health care providers that could have indicated the participants’ level of trust before and after engaging with the mHealth intervention, an important factor in motivation to complete recovery and abstinence from reuse. Future studies should incorporate a larger sample of SMS text messaging for analysis. Additionally, this study could be enhanced by conducting qualitative interviews with participants to understand their motivations as well as facilitators and barriers in using such in-app text messaging services.

Conclusions
In conclusion, by enumerating the types of social support and OUD topic present in the messages exchanged between people in OUD recovery programs and their support staff, this content analysis provides strong evidence for high social support and relapse-prevention needs of people in OUD recovery. Because of addiction and the need for continuous interpersonal support, text-based messages that are instant, reciprocal, and modeled after principles of effective psychological therapies can be the most cost-effective and sustainable solution to providing long-term support for people recovering from OUD.

Conflicts of Interest
PCR is a consultant for Rissana, LLC and PredictView.

References


Abbreviations
- mHealth: mobile health
- MOUD: medication for opioid use disorder
- OUD: opioid use disorder

©Yerina S Ranjit, Warren M Davis, Andrea Fentem, Raven Riordan, Rikki Roscoe, Patricia Cavazos-Rehg. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 10.03.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
eHealth Promoting Stoma Self-care for People With an Elimination Ostomy: Focus Group Study

Igor Soares-Pinto1,2, MSN; Ana Margarida Pinto Braga3, BSN; Isabel Maria Ribeiro Morais Araujo Santos3, MSN; Natália Maria Ribeiro Gomes Ferreira4, BSN; Sandra Cristina da Rocha e Silva5, MSc; Paulo Jorge Alves1,6, MSc, PhD

1Instituto de Ciências da Saúde, Universidade Católica Portuguesa, Porto, Portugal
2Escola Superior de Saúde Norte da Cruz Vermelha Portuguesa, Oliveira de Azeméis, Portugal
3Instituto Português de Oncologia de Coimbra Francisco Gentil, Gabinete de Estomaterapia, Coimbra, Portugal
4Centro Hospitalar Póvoa de Varzim-Vila do Conde, Gabinete de Estomaterapia, Póvoa de Varzim-Vila do Conde, Portugal
5Centro Hospitalar Universitário de Santo António, Gabinete de Estomaterapia, Porto, Portugal
6Centro de Investigação Interdisciplinar em Saúde, Porto, Portugal

Corresponding Author:
Igor Soares-Pinto, MSN
Instituto de Ciências da Saúde
Universidade Católica Portuguesa
Rua de Diogo Botelho, 1327
Porto, 4169-005
Portugal
Phone: 351 22 619 6200
Email: isp.igor@gmail.com

Abstract

Background: The construction of an elimination stoma has a physical, psychological, and social impact on the person. The development of stoma self-care competence contributes to the adaptation to a new health condition and improvement of quality of life. eHealth refers to everything associated with information and communication technology and health care, including telemedicine, mobile health, and health informatics. The use of eHealth platforms by the person with an ostomy, as a digital application that includes websites and mobile phone apps, can bring scientific knowledge and well-informed practices to individuals, families, and communities. It also allows functionalities that enable the person to describe and identify early signs and symptoms and precursors of complications and to be guided to an adequate health response for their problems.

Objective: This study aimed to define the most relevant content and features to promote ostomy self-care integrated into an eHealth platform as a digital app or website to be used by patients for self-management of stoma care.

Methods: We developed a descriptive, exploratory study with a qualitative approach using the focus group methodology, which was oriented to reach a consensus of at least 80%. A convenience sample of 7 participants consisting of stoma therapy nurses was used. The focus group discussion was recorded, and field notes were taken. The focus group meeting was fully transcribed, and a qualitative analysis was performed. The research question was: Which content and features for ostomy self-care promotion should be integrated into an eHealth platform as a digital app or website?

Results: An eHealth platform, which can be a smartphone app or website, for people with ostomy should provide content aimed at promoting self-care, namely in the field of knowledge and self-monitoring, as well as the possibility of interacting with a stoma therapy care nurse.

Conclusions: The stomatherapy nurse has a decisive role in promoting adaptation to life with a stoma, namely through the promotion of stoma self-care. Technological evolution has emerged as a useful tool to enhance nursing interventions and promote self-care competence. The development of an eHealth platform aimed at promoting ostomy self-care should include the capabilities for telehealth and help with decision-making regarding self-monitoring and seeking differentiated care.

(JMIR Hum Factors 2023;10:e39826) doi:10.2196/39826

KEYWORDS
self-care; ostomy; nurses; health education; telemedicine; eHealth
**Introduction**

**Background**

An elimination ostomy is a surgically created opening in the abdominal wall that results in the diversion of feces or urine to the exterior; it may be permanent or temporary [1]. It is estimated that about 1 million people are living with an ostomy, and 100,000 to 130,000 new ostomies are created annually in the United States [2]. In Portugal, there is no clear evidence on the number of people with an elimination ostomy, but it is estimated that there are more than 16,000 ostomies [3]. It is expected that this number will increase since the most likely cause for its construction is colorectal or bladder cancer [4], incidences for which are expected to increase by 60% by 2040 [5].

In addition to the epidemiology, elimination ostomies can be classified according to the anatomical part involved and have different outcomes in patient quality of life and lifestyle. The presence of an ostomy is a life-changing event and has negative implications on various aspects of an individual’s quality of life. It affects those closest to the person, namely family members and caregivers [6]. The person has to deal with not only the diagnosis of a disease and its therapeutic implications but also physical, self-image, and emotional changes, which require necessary adaptations to daily life, including social and professional activities [7]. Therefore, training the person to take care of their ostomy is the responsibility of the clinical team, especially the stomatherapy nurse.

In Portugal, the stomatherapy nurse is a health care professional with advanced skills and knowledge in stoma care. Stomatherapy nurses have a fundamental role in the transition process. Their responsibility and competence to plan, define, implement, and evaluate interventions aim to adapt and modify the person’s reality and prevent complications, thus contributing to an effective training process to promote autonomy and self-care for the stoma [8].

Self-care represents the set of actions that the individual, as well as their family members, performs vis-a-vis their health and well-being, and these actions are directed to keep themselves in shape, maintain good physical and mental health, meet social and psychological needs, prevent illness or accidents, care for minor illnesses and long-term conditions, and maintain health and well-being after an acute illness or hospital discharge [6].

Easy access to information online emerges as a duality for risk and benefit, because if the patient is not able to assess the quality of the information and its suitability for their particular case, the decision-making process can have negative consequences [9]. Particularly in the context of a person with an ostomy, the benefit of this digital communication is providing scientific knowledge and information about well-informed practices to individuals, families, and communities. The person can also describe and identify early signs and symptoms and precursors of complications, as well as receive guidance on the adequate level of health response to their problems, contributing to easier, more appropriate decision-making and leading to reduced costs and economic impact on health [10].

In the scoping review carried out previously by the first author in the context of his doctoral thesis [11], literature was identified that addresses the use of digital tools to support people with an ostomy; however, the studies focused exclusively on communication between the person and the health professional [12].

In a qualitative study conducted in Portugal, also conducted by the first author, that analyzed the nurse’s and patient's perspectives on the promotion of self-care for the stoma, nurses and patients referred to the use of the internet, email, videos, and images as resources. However, they did not refer to an internet site nor a specific tool that had all the resources needed [13]. In addition, we did not identify any platform that provides content aimed at promoting ostomy self-care.

These 2 studies made it possible to identify internationally available resources that promote stoma self-care and its limitations. On the other hand, the studies made it possible to identify the resources used in Portugal and the need to produce valid resources with clinical utility for the promotion of stoma self-care.

Someone undergoing the construction of an elimination stoma is required to have the knowledge and skills to autonomously and effectively manage their new condition. For this purpose, a specific, systematic intervention by nurses in the pre and postoperative periods and after discharge that is directed toward promoting stoma self-care positively influences the path of adaptation to the circumstance of living with an ostomy [14].

Self-care is thus an ongoing process, important for trust and involvement in the new health condition [6]. The development of stoma self-care competence improves the person’s results and is associated with a better quality of life, reducing readmission and complication rates [15].

In the current context of the digital age, the internet has emerged as a source of information, often used by patients and families to obtain information about diseases, treatment options, and care management; for some, it is their first source of information, even before consulting a health professional [16].

As the population interacts with digital technology in almost every aspect of their daily lives, they also expect faster access to answers about their health issues. This is why the intensive use of smartphone and mobile apps offers people new ways to self-assess and monitor symptoms [17].

The use of digital technology in health allows for increased availability of health information, giving people more access, options, and tools to access their health information and communicate with their health team [17].

However, digital literacy does not always match users' health literacy [9]; therefore, easy and quick access to digital health information can have some setbacks. The topics covered, as well as the quality of the content, can vary, and the authors and sources of this information are often unknown. The content can also range from a peer or professional review to personal blogs, opinions, or other people’s experiences [16]. The multiplicity of information can make it difficult and interfere with the selection of the most reliable information needed by the
caregiver and the person with an ostomy in the management of their ostomy [18]. On the other hand, the perception of the quality of health information may change, and the target population may not have the health literacy necessary to assess health information and relate it to their specific condition or case [16]. In fact, it is not enough just to access information; it is also necessary to select, understand, and use it properly and for the intended goals of solving health problems [19].

The term electronic health (or eHealth) refers to health services and information delivered or enhanced through the internet and related technologies [20]; however, this term also can be assumed to be the broadest umbrella encompassing everything that comes with information and communication technology and health care, including telemedicine, mobile health, and health informatics [21], considering eHealth as a tool to promote health or improve health care [22].

In this context, the development of an eHealth platform for support immediately after hospital discharge has emerged; this time period is one of greater vulnerability and when stoma and peristomal skin complications occur. Thus, such an eHealth platform could contribute to reducing the incidence or severity of complications [23]. For this, it is necessary to understand the type of platform and functionalities that would benefit the person with an ostomy.

An eHealth platform may also facilitate the dissemination of guidelines and information regarding the care and self-care of the person with a stoma, strengthening communication as well as the family’s emotional aspects, positively contributing to the transition of the health-disease process [24].

Objectives
The research question that guided this study was the following: Which content and features should be integrated into an eHealth platform as a digital app or website to promote ostomy self-care? Thus, the aim of this study was to contribute to the development of an eHealth platform and define the content and features to be included in this type of platform to promote self-care of the person with an elimination ostomy.

Methods
This was a descriptive, exploratory study with a qualitative approach. We complied with the consolidated criteria for reporting qualitative research (COREQ) with the aim of promoting explicit and comprehensive reporting of interviews and focus groups [25]. The focus group was conducted according to the methodological guidelines defined by Krueger and Casey [26].

Recruitment
A total of 7 participants were involved in the focus group. The number of participants must be sufficient to create discussion, as too large a group may prevent some participants from sharing their ideas within the time available. In this sense, the group size should be from 4 to 12 participants: “the ideal size of a focus group for most noncommercial topics is five to eight participants” [27,28].

The inclusion criterion for the participants was that they had to have advanced competence in stomatherapy as defined by the Portuguese Nursing Board [29].

The sampling process was intentional, seeking to obtain balance in the participants, in order to have professionals with experience in all contexts of care for people for whom construction of an ostomy has been proposed, including the preoperative intervention during the consultation, care of the postoperative inpatient, and follow-up consultation after discharge.

All participants approached agreed to participate in the study. In order to obtain maximum variation in the participants' experiences, differences were considered related to the institutional dynamics and location of the institution, whether in urban or rural areas.

The participants were contacted by email through the Portuguese Association of Stomatherapy Care Nurses (APECE).

In the invitation addressed to the APECE, the motivation for the investigation, its relevance, and the promotion of self-care as a central topic of discussion were explained. No one declined the invitation nor withdrew participation during the study.

Ethical Considerations
Approval was obtained from the ethics committee for health of the Universidade Católica Portuguesa to develop the eHealth platform to promote self-care for people with an elimination stoma (number 141, 246). The confidentiality of all participants was guaranteed, and they were informed that they could withdraw from the study at any time. All participants gave their formal written consent. The location selected for the meetings was at the center of Portugal to facilitate the movement of participants from various parts of the country.

Procedure
Regarding the realization of the focus group, having already had experience in this methodology, the main researcher led the discussion group. The objective was explained, and the exchange of ideas was encouraged, with a second member recording the proceeding and observing the group. This focus group met twice, for approximately 2 hours each session. A script was constructed based on a scoping review on nursing interventions to promote self-care in people with an elimination ostomy [11]. The script also considered the domains that make up the competence of self-care: knowledge, self-surveillance, interpretation, decision-making, execution, and resource management [30]. This script, which was used to conduct the meetings, was based on the questions listed in Table 1.

The literature review carried out and already published [11] was conducted by the first and last authors, being an integral part of the first author's PhD in nursing.

To ensure balance and the achievement of expected results, the group discussion was led by the principal researcher, who ensured equality among group members in terms of intervention in decision-making, while ensuring that all ideas were carefully heard and considered.
Regarding the research assumptions, a consensus of 80% was defined from the outset [31] for all content, asking participants if they agreed with the information and method of exposition, assuming that, if 80% of consensus was not reached, interactive dialogue would continue until 80% agreement was reached.

We also ensured that all issues would be examined in detail and dissenting opinions were duly considered.

Analysis of the information was conducted immediately after the session; no software was used. However, for further analysis, field notes were made throughout the session, to guarantee the consideration and possibility of confirmation that all data were taken into account.

For the data analysis in the first phase, the first 2 authors viewed the recording of the meeting and compared it with the field notes to add aspects not identified in the recording. In the second phase, after transforming the transcribed content into raw data, prepared for grouping, the coding phase was carried out. Categorization followed, which involved the organization and classification of selected text and key points of the transcribed discussions, in context units to form codes. Finally, the interpretation process was concluded, which involved the inferential process that represented the explanation of the codes of the emerging categories and subcategories. Two authors (ISP and AMPB) coded and validated the categories, with approval being obtained from the remaining authors.

The researchers' pre-understanding guided the analysis. All authors held critical discussions, questioning their pre-understanding and theoretical knowledge, to reduce investigator bias throughout the research process.

The authors, who are stoma care nurses, required constant awareness of the need to reduce the risk of portraying their professional experiences and perceptions during all stages of the study.

This pre-understanding was reflected, reconsidered, and examined by the investigators during the process of analyzing and interpreting the data.

To ensure valid and grounded interpretations of the data, we sought to maintain a critical and honest posture through self-reflection.

In the data analysis, we tried to align the theory, objective of the study, data collection, analysis, and results.

Table 1. The focus group’s guiding questions.

<table>
<thead>
<tr>
<th>Question number</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>What is the benefit of using an eHealth platform to promote self-care for an elimination ostomy?</td>
</tr>
<tr>
<td>2</td>
<td>What are the central aspects related to promoting self-care for people with an elimination ostomy that should be considered in constructing an eHealth platform?</td>
</tr>
<tr>
<td>3</td>
<td>What content related to promoting stoma self-care should be part of an eHealth platform intended to promote self-care for bowel elimination stoma?</td>
</tr>
<tr>
<td>4</td>
<td>What strategies can be used to present content on an eHealth platform to promote self-care for the bowel elimination stoma?</td>
</tr>
<tr>
<td>5</td>
<td>In the context of promoting self-care for the bowel elimination stoma, what features should an eHealth platform contain?</td>
</tr>
</tbody>
</table>

Results

The focus group consisted of 7 participants, mostly women, from various parts of the country: 2 from the north, 3 from the center, and 2 from the south of Portugal. We provide their main characteristics in Table 2.

The focus group was formed by the board members of the APECE, an association of recognized relevance in the field of stomatherapy in Portugal.

Regarding the benefits of using an eHealth platform, there was consensus that it responds to an extremely current need, related to the need for remote monitoring. The greatest difficulty in interacting with users and using technologies as support tools in the health area is related to the recognition that there may be population groups with lower possibility of accessing this tool, namely older people.

The results obtained with the focus group were divided into 3 themes: (1) central aspects of self-care, (2) content and methods for promoting self-care, and (3) features of an eHealth platform to promote self-care for bowel elimination stoma.
Table 2. Characteristics of the focus group participants (n=7).

<table>
<thead>
<tr>
<th>Characteristics of the participants</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1 (14)</td>
</tr>
<tr>
<td>Female</td>
<td>6 (86)</td>
</tr>
<tr>
<td><strong>Age (years), mean (range)</strong></td>
<td>45.6 (31-59)</td>
</tr>
<tr>
<td><strong>Length of experience (years), mean (range)</strong></td>
<td></td>
</tr>
<tr>
<td>Nurse</td>
<td>22 (8-33)</td>
</tr>
<tr>
<td>Ostomy care</td>
<td>19 (2-28)</td>
</tr>
<tr>
<td>Stomatherapy consultant</td>
<td>12 (2-20)</td>
</tr>
<tr>
<td><strong>Nurse specialty, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Rehabilitation nursing</td>
<td>1 (14)</td>
</tr>
<tr>
<td>Medical-surgical nursing</td>
<td>4 (57)</td>
</tr>
<tr>
<td>Mental health and psychiatric nursing</td>
<td>1 (14)</td>
</tr>
<tr>
<td>Community nursing</td>
<td>1 (14)</td>
</tr>
<tr>
<td><strong>Had experience with teaching nursing, n (%)</strong></td>
<td>7 (100)</td>
</tr>
<tr>
<td><strong>Postgraduate studies in stomatherapy, n (%)</strong></td>
<td>7 (100)</td>
</tr>
</tbody>
</table>

Central Aspects of Self-care for Bowel Elimination Stoma

Of the domains that make up the competence for stoma self-care, the possible aspects to integrate into an eHealth platform were discussed.

Knowledge and self-surveillance were unanimously considered. An expert said that interpretation and decision-making can also be areas enhanced by the platform, namely identifying complications and the decision to resolve them or seek differentiated help from a stomatherapy nurse:

*A patient looking at the image and comparing it with their stoma can recognize the difference and seek help...* [Nurse 1]

Contents and Methods in Promoting Self-care for the Bowel Elimination Stoma

In the focus group discussion focused on nursing interventions, we tried to define the content and method of promotion to be used in the eHealth platform, as described in Table 3. The content and methods were obtained by consensus greater than 80% (≥6 experts).

It should be noted that issues emerged that, despite not having reached consensus ≥80%, were representative, namely the dietary regimen, with 5 participants suggesting it should be integrated:

*Food is essential for the integration of this new condition into their daily lives.* [Nurse 3]

However, it was concluded that the type of stoma, underlying pathology for the stoma construction, possible neoadjuvant treatments, and person’s associated comorbidities are highly variable factors, which is why it is too complex to establish recommendations in terms of a feeding pattern that considers all the variables listed.

Two experts suggested that sexuality be included:

*The approach to sexuality could be facilitated by the use of this tool...* [Nurse 5]

However, the participants discussed that sexuality could be included, but only some general aspects and information directed toward the available resources should be included.

The group concluded that both diet and sexuality are very relevant topics, but they are not directed toward stoma self-care; they are part of other self-care areas with an impact on integrating this new condition. However, their importance was recognized for the integration of the new condition and promoting quality of life.
Table 3. Content and methods for promoting self-care for the bowel elimination stoma to be integrated into an eHealth platform (n=7).

<table>
<thead>
<tr>
<th>Content</th>
<th>Method</th>
<th>Consensus, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Elimination ostomy: definition, anatomy and physiology, digestive system, types of elimination ostomies, colostomy, ileostomy, urostomy, nephrostomy, collecting devices</td>
<td>Images and text</td>
<td>7 (100)</td>
</tr>
<tr>
<td>2. Pouching system: according to ostomy type, number of pieces, type of fixation (adhesive and mechanical)</td>
<td>Images and text</td>
<td>7 (100)</td>
</tr>
<tr>
<td>3. Stoma site marking</td>
<td>Images and text</td>
<td>6 (8)</td>
</tr>
<tr>
<td>4. Stoma self-care: hygiene, trichotomy/hair removal, device removal, device/plate clipping, device application (single piece and 2 pieces)</td>
<td>Text and video</td>
<td>7 (100)</td>
</tr>
<tr>
<td>5. Stoma and skin self-surveillance</td>
<td>Real images and text</td>
<td>6 (85)</td>
</tr>
<tr>
<td>Stoma observation: standard for normality, stoma complications (stenosis/squeezing, prolapse, ulcer, hemorrhage/blood loss, retraction)</td>
<td>Real images and text</td>
<td>6 (85)</td>
</tr>
<tr>
<td>Skin observation: standard for normality, skin complications (maceration, erythema, hernia, crystal deposits [oxalate/phosphate] on the skin, wounds [ulcer, pressure ulcer, excoriation])</td>
<td>Real images and text</td>
<td>6 (85)</td>
</tr>
<tr>
<td>6. Dietary regimen: colostomy, ileostomy, urostomy</td>
<td>—a</td>
<td>5 (71)</td>
</tr>
<tr>
<td>7. Sexuality</td>
<td>—</td>
<td>5 (71)</td>
</tr>
<tr>
<td>8. Health resources: stomatherapy consultation, pouching system and accessories reimbursement, pouching system and accessories distribution, consume limit on pouching systems and accessories</td>
<td>Text</td>
<td>6 (85)</td>
</tr>
</tbody>
</table>

aDid not reach at least 80% consensus, so methods were not discussed.

Features of an eHealth Platform to Promote Stoma Self-care

Regarding a set of priority features inherent to an eHealth platform, the features listed in Table 4 were understood as essential in a digital tool with the objective of promoting self-care competence for the person with an elimination ostomy.

The possibility of requesting samples of ostomy devices and accessories from the industry through the platform was also discussed; however, there was no consensus from the working group, considering that all materials must be subject to a prior and joint evaluation by the stomatherapy nurse and the person with an ostomy. Before its introduction: it is a very great risk to give the person the possibility to try devices and accessories, without a previous evaluation by the stomatherapy care nurse... [Nurse 7]

The possibility for the platform to include a space for an information sharing forum was also explored. It was understood, however, that due to the difficulty in screening the information transmitted, this item should not be included due to the requirement for a moderator: If there is not a very strict control, less suitable content can be placed that can be read and misused... [Nurse 6]

Table 4. Functions to be integrated in an eHealth platform aimed at promoting stoma self-care.

<table>
<thead>
<tr>
<th>Features</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Telehealth</td>
<td>Interaction via text message, images, or videos with a stomatherapy nurse: The purpose is not to replace face-to-face contact by the health care team but to add an extra resource in case of doubts about or to promote self-care.</td>
</tr>
<tr>
<td>Self-surveillance</td>
<td>Self-surveillance algorithms for the stoma and peristomal skin: The purpose is to help patients identify changes and solutions to problems with the stoma and skin.</td>
</tr>
<tr>
<td>Information about self-care and health resources</td>
<td>Information in the form of videos, images, and text directed at promoting stoma self-care: Make relevant normative and legislative documents available for the person with an ostomy. The purpose is to provide relevant and reliable content and help the patient identify available resources and services.</td>
</tr>
</tbody>
</table>

Discussion

Topics With Consensus

The results show that consensus was reached regarding eHealth tools as current aids with numerous advantages in the context of managing the condition of the person with a stoma. This is corroborated by the literature, which shows the added value of eHealth tools in disease management and health promotion, as well as the enormous potential to promote patient engagement [32].

Considering the 6 domains that integrate the competence of self-care, knowledge, self-surveillance, interpretation, decision-making, execution, and management of health resources [30], a digital platform focused on promoting self-care...
can contribute to the development of knowledge and self-surveillance domains. In addition to these domains, which are also supported by evidence [33], eHealth platforms have also demonstrated impact at the behavioral level, namely by the promotion of skill development and disease self-management [34].

In fact, acquiring knowledge and skills will contribute to the promotion of self-surveillance. Several studies have pointed out that the person with more knowledge about their condition or disease can more easily promote stoma self-care [35].

With regard to the content and methods to be included in an eHealth platform, the group of experts suggested 8 themes to be integrated into the domain of knowledge. Only 6 reached consensus superior to 80%, and the sexuality and dietary themes reached 71% consensus.

Regarding the proposed content of “elimination ostomy” (content 1 in Table 3) and “pouching system” (content 2 in Table 3), the person or caregiver must understand the anatomical aspects to facilitate the understanding of the type of stoma that was constructed, as well as its function and effluent management. These are basic aspects for the promotion of awareness, an aspect that is necessary for promotion of self-care [36]. This content is useful as a starting point for training for self-care, and the availability of this information, even in the initial stages, will contribute to knowledge of the condition. In addition, in later stages, it can be integrated into the decision process, leading to greater involvement in decision-making regarding the devices to be used.

“Stoma site marking” (content 3 in Table 3) is associated with improved quality of life, reduced complications, and better adaptation to the new condition [14]. Marking the stoma site presupposes having a set of assumptions as a basis to identify the best site for its location. These assumptions, in addition to the visible characteristics, which include skin folds, bony prominences, and abdominal morphological aspects, also include the person’s daily activities and clothing [37]. In this way, prior knowledge of these assumptions, the therapeutic route, and its purpose will allow the person to be involved in the decision-making process, facilitate the procedure for marking the stoma site, and reduce preoperative anxiety [1].

To promote “stoma self-care” (content 4 in Table 3), the person needs to receive adequate information as well as support to acquire the skills and resources needed to practice self-care [38].

A person with a stoma is required to have a set of knowledge and skills and the ability to integrate these into the management of stoma care and activities of daily living. The person with an elimination stoma must understand the type of stoma and whether it is temporary or permanent; know the appropriate material to use; know of alternative materials in case the equipment proves to be difficult to use or causes skin problems; have the ability to solve stoma-related problems, including recognizing changes in the stoma, characteristics of elimination, and peristoma skin; and know who to contact if any complications arise [39,40].

Content 5 (Table 3), “stoma and skin self-surveillance,” is also crucial to prevent complications, since approximately 80% of people with an ostomy will experience at least one complication during their lifetime [15] and the absence of complications favorably contributes to the adaptation and integration of the person’s new condition [41].

The person with an ostomy must, therefore, be aware of the appearance of changes and seek help that is appropriate to their needs. People with stomas accompanied by stoma care nurses require 70% less care time, have fewer hospital readmissions due to complications, have one-half the average direct cost of treating complications, and report higher levels of well-being and quality of life and less pain [7].

The use of “health resources” (content 8 in Table 3) to improve the educational process and follow-up of patients through information and communication technologies is rapidly evolving, and nurses play a central and privileged role in the use of these technologies to improve and optimize their interventions with patients [42].

The creation of the stoma involves not only the need for a device but also a new body image that needs to be reconstructed, which is why it is a process that is simultaneously subjective and deeply reflective and requires careful intervention [43].

**Topics Without Consensus**

“Dietary regimen” (content 6 in Table 3) is a very comprehensive topic and specific to each condition, especially for the person with a bowel elimination ostomy.

People with a colostomy have all their small intestines and some functioning colon. The risk of dehydration and malnutrition is therefore not significantly increased. On the other hand, people with an ileostomy do not have a functioning colon and will have varying lengths of functioning small intestine above the ileostomy [44].

Dietary advice is considered an important component of stoma management and is provided by many health care professionals, such as stoma nurses, dietitians, surgeons, gastroenterologists, and other specialist nurses [45].

The literature is not clear about the implications of different stomas and the length of functional intestine on dietary management [45]. Considering that providing recommendations about diet in an eHealth platform is complex, however, this can be a relevant topic to explore and define key points for different types of stomas.

“Sexuality” (content 7 in Table 3) is a highly individualized area of a person’s life and related to all the conditions involved in the health and disease process. Living with an ostomy can have a more noticeable negative impact on aspects of body image and sexuality [1]. If, on the one hand, the person with an ostomy expresses several physiological changes in terms of sexuality, on the other hand, they may also experience psychological problems, such as fear and anxiety related to sexual performance and the possibility of accidents with the device during intimacy. Although the group of researchers considered this an important and integrative item in following up the person with an ostomy, its approach requires an
individualized intervention guided by a health professional who knows methodologies and tools, such as strategies aimed at approaching sensitive topics such as sexuality [46]. Thus, in view of this and considering the unilateral approach of an eHealth platform, its inclusion may have negative repercussions on the intervention process and can only be considered the first phase of addressing the person’s sexual problems—permission, with information directed at the availability of the stomatherapy nurse to address this issue and find solutions together [46].

Features of an eHealth Platform

Regarding the features of the eHealth platform, rapid advances in technology and internet access have become not only a viable way to carry out educational interventions but also a platform that can be widely disseminated and implemented. In addition, internet interventions and programs that can be disseminated through the internet allow program content to be standardized, targeted to specific ages and developmental stages, and easily updated [47].

Faced with the challenge of an ostomy, the task of empowering the person with an ostomy is up to health professionals, namely stomatherapy nurses, providing them with knowledge and skills to manage their self-care. This task is usually carried out face to face in a programmed teaching context; however, the use of digital technologies favors more flexible access to information, allowing the person to search for and process it at their own pace [48].

The use of telehealth after hospital discharge is effective in improving satisfaction with care, reducing stoma complications, improving self-care competence, and increasing the patient’s self-confidence in dealing with the ostomy. Although performed at a distance, follow-up has become an extremely important factor for better adaptation to the ostomy [49]. There are, however, barriers to the use of digital information, which include lack of access to the internet and a low level of digital literacy in health [17].

With regard to the use of surveillance algorithms for the stoma and peristomal skin, continuous monitoring is one of the several areas of intervention by nurses that contribute to managing the new health condition of the person with an ostomy [1].

Limitations

Despite the relevance of the results obtained, this study has some limitations. The discussions were held only once in the group, although 2 meetings were held, and whether the participants’ perception remained constant regarding the final results was not evaluated.

The literature review carried out and directed at nursing interventions to promote self-care was shared with the participants only on the day of the meeting, with no time for discussion of the information and confrontation with each person’s experience and taking the limitations of their own review into account [11].

Obtaining consensus has inherent limitations, namely not generating new knowledge but reflecting the opinion of experts. The results obtained could be different in larger groups.

Conclusion

People with an elimination ostomy have their perspective of life altered, due to fear and doubts about the ostomy and devices to be used. Furthermore, people must manage various changes that occur in their daily lives, namely eating habits, hygiene, physical activity, and professional activity, as well as many other aspects necessary for their adaptation, which will have implications on their lifestyles [41,50].

Regardless of the approach, face-to-face or through eHealth platforms, the stomatherapy nurse has a wide field of action, interventions, and strategies to care for people with an ostomy and can enhance the person’s adaptation to their new condition.

This study made it possible to define a set of content and key features to be included in an eHealth platform focused on promoting self-care.

It is necessary to evaluate these new ways of communicating with patients with an ostomy and caregivers, which can facilitate the promotion of self-care and autonomy and can enhance social reintegration, through specialized support and monitoring.

Despite so many challenges and the complexity of the situation, when the person with an ostomy is adapted with adequate support and resources, it is possible to live an active and quality life [51].

Acknowledgments

The authors thank the Portuguese Stomatherapy Care Nurses Association (Associação Portuguesa de Enfermeiros de Cuidados em Estomaterapia-APECE) for their contribution. Additionally, the authors extend their appreciation to SPCare for their support and willingness to support the study.

Conflicts of Interest

None declared.

References


27. Beyea S, Nicoll L. Learn more using focus groups. AORN J 2000 Apr;71(4):897-900 [FREE Full text] [doi: 10.1016/s0001-2092(00)62276-x] [Medline: 10806545]


Abbreviations

APECE: Portuguese Association of Stomatherapy Care Nurses
COREQ: consolidated criteria for reporting qualitative research
Experiences of Using an Electronic Health Tool Among Health Care Professionals Involved in Chronic Obstructive Pulmonary Disease Management: Qualitative Analysis

André Nyberg\(^1\), PhD; Anna Sondell\(^1\), PhD; Sara Lundell\(^1\), PhD; Sarah Marklund\(^1\), MSc, PhD; Malin Tistad\(^2\), PhD; Karin Wadell\(^1\), PhD

\(^1\)Department of Community Medicine and Rehabilitation, Section of Physiotherapy, Umeå University, Umeå, Sweden
\(^2\)School of Education, Health and Social Studies, Dalarna University, Falun, Sweden

Corresponding Author:
André Nyberg, PhD
Department of Community Medicine and Rehabilitation
Section of Physiotherapy
Umeå University
Universitetstorget 4
Umeå, 901 87
Sweden
Phone: 46 907866639
Email: andre.nyberg@umu.se

Abstract

Background: Chronic obstructive pulmonary disease (COPD) is one of the most common and deadliest chronic diseases of the 21st century. eHealth tools are seen as a promising way of supporting health care professionals in providing evidence-based COPD care, for example, by reinforcing information and interventions provided to the patients and providing easier access and support to the health care professional themselves. Still, knowledge is scarce on the experience of using eHealth tools from the perspective of the health care professional involved in COPD management.

Objective: The study explored the experiences of using an eHealth tool among health care professionals that worked with patients with COPD in their daily clinical practice.

Methods: This exploratory qualitative study is part of a process evaluation in a parallel group, controlled, pragmatic pilot trial. Semistructured interviews were performed with 10 health care professionals 3 and 12 months after getting access to an eHealth tool, the COPD Web. The COPD Web, developed using cocreation, is an interactive web-based platform that aims to help health care professionals provide health-promoting strategies. Data from the interviews were analyzed using qualitative content analysis with an inductive approach.

Results: The main results reflected health care professionals’ experiences in 3 categories: receiving competence support and adjusting practice, improving quality of care, and efforts required for implementation. These categories highlighted that using an eHealth tool such as the COPD Web was experienced to provide knowledge support for health care professionals that led to adaptation and facilitation of working procedures and person-centered care. Taken together, these changes were perceived to improve the quality of care through enhanced patient contact and encouragement of interprofessional collaboration. In addition, health care professionals expressed that patients using the COPD Web were better equipped to tackle their disease and adhered better to provided treatment, increasing their self-management ability. However, structural and external barriers bar the successful implementation of an eHealth tool in daily praxis.

Conclusions: This study is among the first to explore experiences of using an eHealth tool among health care professionals involved in COPD management. Our novel findings highlight that using an eHealth tool such as the COPD Web may improve the quality of care for patients with COPD (eg, by providing knowledge support for health care professionals and adapting and facilitating working procedures). Our results also indicate that an eHealth tool fosters collaborative interactions between patients and health care professionals, which explains why eHealth is a valuable means of encouraging well-informed and autonomous patients. However, structural and external barriers requiring time, support, and education must be addressed to ensure that an eHealth tool can be successfully implemented in daily praxis.
**Introduction**

Chronic obstructive pulmonary disease (COPD) is one of the most common chronic diseases of the 21st century and a leading cause of chronic morbidity worldwide [1]. COPD is typically treated and managed with pharmaceutical and nonpharmacological therapies in primary, secondary, or tertiary care [1-4]. Although nonpharmacological treatments, such as pulmonary rehabilitation and self-management interventions, are considered vital components of COPD management [2,5], we know today that several barriers exist that result in low access, uptake, and completion rates [6-8]. Therefore, overcoming these barriers and finding new and alternative strategies to facilitate evidence-based care in COPD management are highly warranted [9,10].

eHealth tools represent a promising way of delivering health services in COPD management [11-13]. The use of eHealth tools includes, but is not limited to, intervening, educating, and keeping track of a person’s health, resulting in several clinically relevant health benefits among patients with COPD [11,12,14-16]. For example, our group previously found that access to a web-based platform increased self-reported physical activity levels, COPD-specific knowledge, and altered disease management strategies among patients with COPD in primary care [17]. However, we also found that the use of the eHealth tool varied profoundly between patients, and the vast majority mainly used the platform at the beginning stages of their treatment [17]. Furthermore, motivation, comfort with information technology tools, and level of health literacy were identified as vital explanatory factors affecting usage of the eHealth tool over time [18], findings that are supported by a recent qualitative systematic review that determined the perception of eHealth among over 300 patients with COPD across 19 individual studies [19]. However, besides motivation and comfort with information technology tools, other factors, such as access to 1-to-1 contact with health care professionals, were also critical for encouraging use of eHealth tools among patients with COPD [19]. Regarding the latter, van Zelst et al [20] recently demonstrated up to a 3-fold increase in eHealth tool usage among patients with COPD if the tool was used together with health care professionals compared with those who used the eHealth tool independently. This indicates the vital role of the health care provider in supporting the use of eHealth tools among patients with COPD.

Importantly, these tools are also accessible and potentially relevant for users other than patients, such as health care professionals and informal caregivers [21,22]. Recently, eHealth tools have been put forward as a viable alternative supporting health care professionals in providing evidence-based care, for example, by reinforcing information and interventions provided to the patients and providing easier access and support to the health care professional themselves [23,24]. Furthermore, eHealth tools are considered a promising way for health care professionals to interact with and support patients and their families at a distance [25,26], and they may improve patient-related outcomes and health care utilization by providing self-management support for the patient and decision support for the health care professional [26]. Besides, attitudes toward using eHealth tools among health care professionals are generally positive. In a recent global survey among 1091 health care workers, 4 out of 5 health care professionals thought that using eHealth tools can reduce workload and save time for the clinician [27]. Yet, despite the potential benefits of eHealth tools for health care professionals, and the generally positive attitude toward using eHealth tools, knowledge is scarce on the practical experience of using eHealth tools from the perspective of the health care professional, specifically the health care professional involved in COPD management [24,28,29]. In addition, a need for further qualitative research was warranted in a recent meta-synthesis to understand the key ingredients that will facilitate a positive user experience of eHealth among health care professionals [29]. About the latter, designing eHealth tools using co-creation or participatory methods that engage the end users in the development and design of the eHealth tool is recommended but not often used within COPD research. Therefore, using an explorative qualitative design, this study explored the experiences of using an eHealth tool that was designed using a cocreative process [24] among health care professionals working with patients with COPD in a primary care setting.

**Methods**

**Study Design**

This exploratory qualitative study is part of a process evaluation in a parallel group (1:1 allocation) controlled pragmatic pilot trial [10,17], reported per the Consolidated Criteria for Reporting Qualitative Research (COREQ) guidelines [30]. The study was registered at ClinicalTrials.gov (NCT02696187).

**Ethical Considerations**

Ethical approval was given by the Regional Ethical Board, Umeå University, Umeå, Sweden (Dnr: 2014-319-31, 2015-457-32). In addition, written informed consent was obtained from each health care professional before their enrollment in the study. Study data are not anonymous but were deidentified. No compensation was provided to study participants.
Setting and Sample
A total of 10 health care professionals at 5 publicly funded primary health care centers (2 situated in northern Sweden and 3 in central Sweden) were invited to participate using convenience sampling. All 10 health care professionals accepted. Primary care was targeted, because it is where the vast majority of patients with COPD in Sweden are treated [31,32]. The senior manager at each primary care unit assisted in identifying health professionals eligible for participation in the study. Telephone calls or emails were used to approach potentially eligible health care professionals at the included centers to participate in the planned study. To qualify for inclusion, health care professionals, independent of profession (eg, nurses, physicians, physiotherapists, occupational therapists, or dietitians), should meet patients with COPD in their daily clinical practice and be willing to use an eHealth tool, the COPD Web, as part of their clinical praxis for at least three months. As part of the process evaluation, interviews were performed with health care professionals involved in COPD management at 3 and 12 months, the latter to capture the longitudinal long-term experience of using the eHealth tool.

The eHealth Tool
The COPD Web is an interactive web-based platform that was cocreated with patients with COPD, their relatives, health care professionals, and researchers. The content of the COPD Web was in line with the nonpharmacological health promotion interventions recommended by the Swedish National Board of Health and Welfare’s national guidelines for COPD management [33]. The COPD Web consisted of 3 main sections, 1 directed at patients with COPD, 1 at their relatives, and another at health care professionals [24]. The COPD Web’s development and design and the experience and effects of using the COPD Web among patients with COPD have all been extensively described elsewhere [10,17,24].

The health care professional section of the COPD Web aims to support evidence-based care for patients with COPD, specifically self-management strategies. The section included factual texts, pictures, videos, and recommended and validated evaluation and screening tools [10]. An overview of the COPD Web’s current content, specifically, the section for health care professionals, is shown in Figure 1. Data on use of the COPD Web were gathered during the initial 3 months. Health care professionals made on average 15 (SD 19) log-ins to the COPD Web and spent 15 (SD 21) minutes on the site per log-in. Across the 10 health care professionals, the COPD Web was introduced to 102 patients with COPD during the 3 months.
Figure 1. Overview of the content of the COPD Web sections for health care professionals. My page was a specific section of the COPD Web that became available when creating an account. In the My page section, the user could change settings, find contact information for their primary care center that was selected when creating the account, and find an overview of the “favorites” sections of the COPD Web. About the latter, each page of the COPD Web could be saved as a “favorite” and the user could have an overview of the sections that had been saved on the My page section. The purpose was to provide a quick and direct access to the specific content on the site being important for each user. COPD: chronic obstructive pulmonary disease.

Introduction of the eHealth Tool
All health care professionals were given a 1-2-hour theoretical and practical face-to-face introduction to the COPD Web. The information provided was predetermined and similar across health care centers and independent of the profession of the health care professional. The introduction included information on the design, development, and purpose of the COPD Web. As part of the use of the COPD Web, the health care professionals were also instructed on how to introduce the COPD Web, using a prespecified routine (Textbox 1), to all patients with COPD that they met during the initial 3-month period. We went through all the steps highlighted in Textbox 1 with the health care professionals as part of the 1-2-h education, with the health care professionals also navigating the site. Besides the introduction of the COPD Web and in line with the pragmatic approach of the study, the health care professionals were free to use, or not use, or to adapt the use of the COPD Web as they deemed suitable or appropriate for each patient. No extra resources were provided to the primary care units, as health care professionals used the COPD Web as a part of their regular work practice [10].

Textbox 1. Routine introduction of the COPD Web to patients by health care professionals. Reproduced, with permission, from [10]. COPD: chronic obstructive pulmonary disease.

- Registration and creation of an account to allow the patient to use the COPD Web.
- Introducing the website structure, the content of the main menus, and the functions of the website; for example, how to enlarge or reduce text, listen to the text, or bookmark information of particular interest.
- Introducing the “physical activity and exercise training” section to the patient. The health care professional will discuss the importance of physical activity/exercise training, point out the films with muscle strengthening exercises, and the page for registering physical activity (steps) with automated feedback.
- Introduction of 2 to 4 additional topics on the website of particular interest for the specific patient in question.
- Topics of specific interest for the patient will be noted on a leaflet with information about the COPD Web. The patient will receive the flyer and a card with the COPD Web’s URL address, username, and password.
Research Team

The research group consisted of physiotherapists with different preunderstandings and insider and outsider perspectives of the eHealth tool. Two researchers, AN (PhD, male, 32 years) and MT (PhD, female, 43 years), conducted the interviews separately. Both interviewers were employed as postdoctoral researchers at the Department of Community Medicine and Rehabilitation, the section of Physiotherapy at Umeå University, Sweden, at the time of the study. Before the trial commenced, AN had conducted more than 20 interviews without specific prior training, while MT had conducted over 30 interviews under supervision during a previous postdoctoral employment. Before the interviews, there was no relationship between AN/MT and the health care professionals enrolled in the study. However, health care professionals knew that AN and MT had been involved in developing the COPD Web.

Process of Data Generation

Semistructured individual interviews (except 1 made in pairs) with open-ended questions were conducted by AN or MT. Overall, 10 health care professionals (4 women), including 5 nurses, 2 physiotherapists, 1 dietician, 1 occupational therapist, and 1 physician, with a mean age of 50 (SD 11) years and 25 (SD 11) years of work experience, participated in the interviews; 5 health care professionals accepted interviews at 12 months. Reasons for declining an interview at 12 months included not using the eHealth tool (n=4) and being retired (n=1). Interviews were conducted in the health care professionals’ workplace at 3 months and over the telephone 12 months after receiving access to the COPD Web. In all interviews, only the interviewer and the health care professional were present. To ensure health care professionals’ privacy, all names were changed to pseudonyms during the start of the analysis so that only interviewers knew participants’ real names. No immediate callbacks on the interviews were conducted (ie, for potential amendments or additional questions). Still, exciting or unexpected topics raised during the interviews were discussed and used to guide follow-up questions during the following interviews.

The 3-month interviews ranged between 12 and 67 minutes (mean 39 minutes), while the 12-month interviews ranged between 10 and 29 minutes (mean 19 minutes). The interviews were structured by an interview guide (Multimedia Appendix 1), including questions about the professionals’ experiences using the eHealth tool, its applicability and usefulness, knowledge support and added value, and what they thought was missing from the eHealth tool when working with patients with COPD in primary care. For example, the first question was: “Tell me about if/how you have used the COPD Web (during this time)?” Participants were encouraged to speak freely in responding to the questions, and the interviews proceeded as conversations. Transcripts were not returned to health care professionals for comment or correction, and health care professionals were not engaged to provide feedback on the findings. All participants were assigned a pseudonym for transcription proofreading (used when quoting in the “Results” section).

Data Analysis

Data from the interviews were analyzed using qualitative content analysis with an inductive approach, according to Graneheim et al [34,35]. Qualitative content analysis involves a stepwise, systematic analysis and a process of interpretation that focuses on similarities and differences found in the material, resulting in data organization into subcategories, categories, and potentially themes. This procedure is considered an appropriate method for illuminating health care professionals’ experiences of a complex phenomenon in a structured manner and is useful when dealing with already gathered qualitative data [34]. The unit of analysis was all interviews. One author (AS) who had not previously been engaged in the development of the eHealth tool or involved in data collection was chiefly responsible for data analysis. First, interviews were read through several times (with the assistance of audio recordings for auditory cues). Next, the transcripts’ content was divided into meaning units consisting of constellations of words and statements with the same meanings. Meaning units were then condensed and coded using Open Code software 4.03 [36] by one author (AS), with independently parallel coding conducted by 2 authors (KW and SM) in 2 interviews. Based on similarities and differences between codes, preliminary subcategories were clustered, abstracted, and merged into categories. The interpretive process was made in several steps and the analytical process involved a back-and-forth movement between the whole and parts of the texts. Through the analysis process, triangulation between researchers with different backgrounds was used to attain higher credibility [34]. All authors were involved in creating subcategories and categories, and changes were made until consensus was achieved. Trustworthiness was sought, for example, by all coauthors’ participation in several steps of the analysis, and the authors’ complementary competencies and perspectives were of great importance during analysis. In this study, the authors were all physiotherapists (AN, AS, MT, SL, SM, and KW) with clinical expertise in COPD (AS, SL, SM, and KW), specialist competence in COPD and exercise training (AN and KW), and scientific expertise in COPD (SL, KW, AN, SM, and MT), in eHealth (AN, SL, MT, SM, and KW), in exercise training/rehabilitation (AN, AS, MT, SM, and KW), and in qualitative research (AS, SL, MT, and SM). During discussions pertaining to data analysis, researchers critically reflected upon their prior understanding.

Results

Overview of Categories

The analysis resulted in 9 subcategories, grouped into 3 categories: receiving competence support and adjusting practice, improving the quality of care and efforts required for implementation, and representing the experiences of using an eHealth tool among health care professionals working with COPD (Figure 2).

https://humanfactors.jmir.org/2023/1/e43269

JMIR Hum Factors 2023 | vol. 10 | e43269 | p.552

(page number not for citation purposes)
Receiving Competence Support and Adjusting Practice

Overview
The category addresses how the COPD Web provided competence support for health care professionals and patients, and how the work was adapted accordingly.

Professional Knowledge Support
The health care professionals communicated that the COPD Web was a complementing pedagogical and more extensive toolbox that could facilitate the patient meeting. They emphasized the COPD Web’s advantage as a concentrated, evidence-based, cutting-edge, and unified knowledge bank. Using the COPD Web led them to receive a higher and broader level of competence regarding COPD and provided support in patient education, which improved their ability to offer the patients more knowledge. In addition, the fact that the COPD Web provided patients with similar information as the health care professionals was perceived as an advantage.

The COPD Web was perceived to be a support when patients asked questions outside the health care professional’s specific area of expertise, which was considered reassuring. It was also considered more illustrative and spontaneous than using, for example, brochures in patient care.

It’s been a tool I’ve used spontaneously; instead of taking a textbook or chart or compendium, I’ve resorted to the COPD Web without really thinking about it. [Health care professional 1]

Health care professionals experienced increased knowledge about physical activity and exercise, which facilitated prescription of exercise to patients. The health care professionals also expressed that they had gained increased knowledge about using scales and tests in assessing and evaluating aspects of patient health.

I knew nothing about them before, and I think that it’s been quite good to learn a bit about this Borg scale and how it can be used for cardio and strength training. [Health care professional 2]

In addition to being a knowledge support in preparing for patient encounters, health care professionals suggested that the COPD Web would be specifically advantageous when introducing new personnel.

Self-management Support for Patients
Health care professionals emphasized that the COPD Web contained excellent self-management support for the patients (eg, practical tips for managing daily activities and efficient strategies for avoiding exacerating symptoms). The health care professionals expressed that the patients who were more affected during their everyday lives were also more receptive to information and emphasized that well-informed patients with a higher level of knowledge about their disease also adhered better to the prescribed treatment. As a result of using the COPD Web, the health care professionals expressed that they now talked more about what the patient themselves can do in the event of deterioration and that they finally received positive responses regarding self-management from patients.

That’s where things changed! Goodness gracious. Because that’s who I was on the phone with. That’s where it clicked. How to take care of yourself, how to be active and the importance of both treatment and self-care and not to overexert yourself. Really, it’s a revolution. And it’s actually fun. [Health care professional 1]

According to the health care professionals, patients and relatives had shown great interest in the COPD Web, even though interest in the COPD Web was perceived to vary between patients. Younger patients and patients more affected by their disease in their everyday lives were perceived to be the most frequent users of the COPD Web. In addition, health care professionals described that patients now had more time to learn new and essential things concerning their disease thanks to the COPD Web, which could lead to more questions during appointments and better adherence to treatment. Furthermore, patients had expressed lessons they had learned using the COPD Web (eg, finally understanding their disease, feeling it is okay to live with COPD, and not having the same anxiety about it getting worse). Improvements in physical function and increased motivation by following their physical activity over time were also said by patients as positive consequences of using the COPD Web:

Don’t underestimate the part about competition and wanting to beat your own record and...even if you’re a bit old and unwell and all, many people have a competitive streak and might think it’s fun to compare and write up...but it can still be a bit of fun and...well, I think it can be fun. [Health care professional 3]
Health care professionals emphasized that “patient stories” (the section of the COPD Web containing short video interviews with patients with COPD) were pedagogical in-patient education. Seeing other patients’ experiences and solutions provided valuable knowledge and support for patients:

> If I’ve learned anything in my years in health care, I’ve seen that many times it can be a positive teaching method. I think, a patient telling something to another patient. So it’s not just me, the health care professional, who is the storyteller. [Health care professional 4]

### Adapted Working Procedure

The COPD Web was perceived as logical, easy to navigate, and feasible to adapt to the patients’ needs. Further, it could help provide structure when meeting patients and include materials they could use in education classes of patients with COPD. Health care professionals expressed that they, in dialog, worked practically and reflected with the patient in connection with the COPD Web content. Specifically, videos were experienced as facilitating communication with patients and as making it easier for patients to absorb information.

> Yes, the films are good because they come in several different ways, visual and auditory and maybe textual as well. They come to the patient via several channels, so to speak, making it very, very strong. [Health care professional 1]

In addition, it was expressed that because of the COPD Web, finding suitable activities/exercises at the right level for different patients was easier. Assessments of physical function contained in the COPD Web also facilitated the prescription of exercises at appropriate levels. Furthermore, some health care professionals expressed that they had started to send out the standardized questionnaire with the usual invitation for the next checkup, as well as asking the patient to fill in “my COPD profile” in advance, leading to a more thorough consultation. As a result, health care professionals emphasized that they could meet the patient’s needs more directly and be better prepared than they had been before using the eHealth tool. Most health care professionals expressed that they would continue to use the COPD Web and work more with it in the future. One participant expressed that it was undesirable to work without the COPD Web:

> Well, I couldn’t work without it. I just want to show it to everyone. So everyone can use it. All the doctors, everyone...so that they understand. [Health care professional 3]

However, for some health care workers, the COPD Web did not affect the ways in which they worked or their dialog with patients. In addition, after the 12-month intervention period, some health care professionals mentioned that they did not use the COPD Web as often and were not as structured as they had been during the initial months.

### Improving the Quality of Care

#### Overview

This category refers to the importance of enabling a person-centered usage or personalization of an eHealth tool, that is, the importance of the eHealth tool to enable an added value for the health care professional (of using the tool) which taken together can improve COPD management.

#### Person-centered Usage

Health care professionals emphasized the importance of a person-centered usage of the COPD Web. They described meeting patients at different stages of the disease, and the importance of being able to individualize the information given to the patient. They thought that the design of the COPD Web enabled this as the various sections of the COPD Web were applicable to patients of different ages and stages of the disease. However, individuals who considered themselves “healthy” indicated that they did not always recognize themselves in the information. Therefore, these health care professionals requested even more diverse information on the disease for different stages of disease severity. They pointed out that the information given to patients had to be individually tailored, dependent on a patient’s problems, needs, abilities, prerequisites, and resources. Furthermore, it was important that the information provided sounded familiar, and that patients could see themselves in what was presented.

> I think...the important thing is that it resonates. Even if it’s not that particular activity, there’s something you can relate to. So I think it’s good and illustrative. Then you can fill it in yourself. You can’t have everything on film. Nah. [Health care professional 5]

The health care professionals described that (when meeting a patient) the information on the COPD Web was chosen based on the patients’ individual needs and how they used the COPD Web was adapted to the patient in front of them. They further expressed that individually tailored (physical) meetings were still crucial when motivating patients to self-manage their health, as specific strategies are challenging to illustrate on video and in writing (eg, how to divide an activity into partial tasks for the purposes of energy conservation).

For future development, it was recommended that the COPD Web could be further developed regarding its content nutrition and emotional support.

> Maybe you can clarify something on the front page...COPD is a varied disease and some need more help than others, and this page is adapted for the whole range, so some of what is written apply to those who are very sick. [Health care professional 6]

#### Enhanced Patient Contact

The health care professionals described that using the COPD Web had initially resulted in more extended visits. However, they pointed out that conveying information must be allowed to take time—time that was considered well-invested, as they were confident that the extra time they spent using the COPD Web would pay off in the future through healthier patients.
It contains more information than what I’ve given before, so it’s a qualitatively better meeting than the 45 minutes I gave before. Now, I spend an hour. And I think...I’m quite sure I’ll it will pay of next time. [Health care professional 1]

The health care professionals that used the COPD Web frequently described fewer emergency visits for the patients at the health care centers in which they worked, because they had changed their working procedures to be more aimed at prevention. In addition, the COPD Web facilitated telephone counseling and led to more telephone follow-ups instead of physical visits, which saved time. The health care professionals that used the COPD Web more frequently were convinced of its benefit and security for the patient. After using the COPD Web, some health care professionals described they wanted to reintroduce annual visits for patients that they had not hitherto prioritized, due to lack of time.

Inspiring Interprofessional Collaboration

Health care professionals emphasized that the COPD Web could contribute to collegial and interprofessional collaboration at the health care center, and within the county council, and that this collaboration would be important in the future. They expressed that they now made more contacts with other health care professions working with patients with COPD at their health care center. Different health care professionals used parts of the COPD Web differently and discussed different topics with the patients.

But one of its strengths is that it is so broad, there is so much variety. Different skills and different parts. [Health care professional 3]

Furthermore, although interprofessional collaboration was considered crucial, the health care professionals pointed out that the content of the COPD Web highlighted physiotherapists’ vital role in COPD management specifically, and that physiotherapists needed to take more responsibility for COPD management in the future. In addition, it was addressed that the occupational therapist has a vital role regarding, for example, energy conservation techniques, and should therefore be involved more. They further expressed a wish to extend the content of the COPD Web to include hospital care, home care, and group treatments for COPD. Furthermore, health care professionals pointed out that some of the COPD Web’s information could be helpful to other patient groups, such as those with heart failure.

Efforts Required for Implementation

Overview

This category presents the process of learning to use the COPD Web among health care professionals, what they found to be necessary for the tool’s implementation, and barriers they experienced when using the COPD Web.

A Learning Process

Health care professionals expressed that learning and becoming familiar with the COPD Web took some time, and thus, more time and information on how to use the tool initially was warranted. They especially needed to think through how to use the COPD Web and familiarize themselves with the information they wanted to show patients.

I’ve just browsed and looked around at what there is under different things, so that I can find it later. [Health care professional 3]

It was further emphasized that it was important that the COPD Web operated smoothly and that it did not take too long to find what was needed when a patient was present in the examination room. Health care professionals pointed out that a delay when starting videos and too-long videos could lead to lost focus and disrupt the meeting with the patient. At 12 months, the COPD Web seemed more integrated into daily work among those who continued to use the eHealth tool. Health care professionals became more accustomed to the tool the more often they used it during patient encounters.

Need for Implementation Support

Health care professionals expressed several matters they considered essential when implementing a tool like the COPD Web in daily clinical practice. First, they pointed out that changing routines and learning a new way of working take time. Thus, implementing new procedures in health care—specifically in primary health care with its complex and varied assignments—could be associated with resistance, especially when new methods initially take more time. Second, they further emphasized that health care professionals need to see that the tool is beneficial not only for the person with COPD but also for the health care professionals themselves during their everyday work.

You have to be sure that it benefits the staff. And if you’re not sure, then you have a problem. I guess that applies to everything you introduce in terms of working methods. [Health care professional 1]

Third, they emphasized that support from management is essential and the importance of proceeding cautiously in the event of novel work procedures. Besides, the significance of collegial support when introducing something new was emphasized. Fourth, a good introduction and education should not be neglected, including the opportunity to try out and use the tool in practice. Lastly, the health care professionals expressed the importance of using role models when implementing an eHealth tool such as the COPD Web (eg, intercollegial spread in health care centers where it is already being used and involving health care professionals who have already used it).

Perhaps spread it among those of us involved, in small groups to start with - don’t overreach but spread it small. [Health care professional 1]

Perceived Barriers for Usage

Health care professionals perceived that a barrier to getting started with the COPD Web was when the work procedure was not anchored in primary health care centers in advance. In addition, they described not feeling involved in the decision to start using the COPD Web.

I think it was a bit crazy that we never talked about it here before, so it just sort of came and...nah, I don’t

[Health care professional 1]
The health care professionals also felt the need for certain prerequisites to be met before they could use the COPD Web in their clinical work (eg, a computer in the consulting room or that the patient was not too sick during the visit). A lack of time and needing to prioritize other tasks were also mentioned as obstacles to using the COPD Web. It was pointed out that if the COPD Web did not feel like a natural part of the patient visit, it was not used at all. The biggest obstacle to not using the COPD Web at the health care centers was patient related. The advantages of the patient using the tool at home were emphasized, but it was more difficult when they had no computer at home, or their computer skills were deemed too low. Computer skills were perceived to vary between patients, where greater computer skills seemed to be related not to gender, but to younger age and higher education. Although exceptions to this rule were sometimes apparent.

There’s...I met an 84-year-old man who showed me his computer driving licence, which he took 20 years ago, and he had no problems. And then there were sixty-year-olds who are not at all used to it. [Health care professional 2]

Discussion

Principal Findings

This study explored the experiences of using an eHealth tool, the COPD Web, that was developed using a cocreative process among health care professionals working with COPD in primary care. The main results reflected study participants’ experiences in the following 3 categories: receiving competence support and adjusting practice, improving quality of care, and efforts required for implementation. The uniqueness of this study is the longitudinal design with both 3 and 12 months of follow-up, the use of an eHealth tool that was designed in a cocreative process, and the pragmatic trial design in which the qualitative analysis was part of a process evaluation. About the latter, the eHealth tool was used as part of daily clinical practice and except for a 1-2-h initial education no additional support was provided to the health care professionals, and they were free to use or not use the tool as they preferred. The findings of this study highlight that using an eHealth tool such as the COPD Web was experienced as providing knowledge support for health care professionals, leading to adaptation and facilitation of work procedures and person-centered care, enhanced patient contact, and encouragement of interprofessional collaboration. Taken together, use of the eHealth tool was experienced to improve the quality of care provided to the patient. Health care professionals also expressed that patients using the COPD Web were more well-informed and better equipped to manage their disease. The patients also adhered better to treatment, thus improving their self-management ability. The latter is of utmost importance as self-management is one of the cornerstones of successful COPD management for which eHealth tools may play a vital role. Furthermore, the 3- and 12-month data collection enabled novel insights into how the use of an eHealth tool could change over time, and that among those who continued to use the tool, it was now an integrated part of their daily clinical practice. Still, despite the positive experience of using an eHealth tool among health care professionals involved in COPD management, structural and external barriers requiring time, support, and education must be addressed to ensure that an eHealth tool can be successfully implemented in daily praxis.

Interpretation of Findings

Numerous studies have investigated eHealth tools’ experiences among patients with COPD. However, to our knowledge, this study is among only a few that has explored the experiences of using eHealth tools among health care professionals involved in COPD management. Our results support previous research that claims that eHealth fosters collaborative interactions between patients and health care professionals, which explains why eHealth is a valuable means of encouraging well-informed and autonomous patients. For example, in the category “improving the quality of care,” health care workers repeatedly expressed that using the COPD Web with their patients had resulted in better and more qualitative visits and more well-informed patients that were better equipped to handle their disease and adhered better to their treatment—increasing the patient’s ability to self-manage their disease. In addition, health care professionals emphasized the importance of “person-centered usage” of the COPD Web and the idea that the tool could be used to personalize treatment. For example, it was expressed that different sections and subsections of the COPD Web were necessary for patients at different ages and stages of the disease, and that the content could be adapted, and individualized, depending on which patient sat in front of them, thus highlighting the importance of individualization or “person-centered usage” of an eHealth tool as a potential key ingredient for a positive user experience among health care professionals. Furthermore, similar to our findings, previous meta-analyses have demonstrated a positive relationship between an autonomy-supportive health care climate and the personalization of eHealth intervention contents, successful self-management, and behavior change.

Moreover, we have previously reported that a digital COPD education program could be used to increase objective measures of COPD-specific knowledge among health care professionals involved in COPD management. Although objective measures of COPD-specific knowledge were not obtained in this qualitative study, in the category “receiving competence support,” health care professionals expressed that the COPD Web offered a unified knowledge bank, which led them to receive a higher and broader level of competence regarding COPD. Specifically, health care professionals expressed that their increased knowledge about physical activity and exercise facilitated the prescription of exercise to the patients, which is highly important considering the benefits of physical activity and exercise training in COPD. Furthermore, inadequate professional competence, lack of person-centeredness, and limited access to evidence-based care have been identified as essential obstacles and barriers to prescribing exercise and physical activity interventions to patients. Therefore, the notion that using an eHealth tool may increase competence,
patient-centeredness, and facilitate exercise prescription is an important finding supporting the relevance of eHealth as a means of improving quality of care.

Health care professionals also perceived that the COPD Web contributed to collegial and interprofessional collaboration. For example, as expressed in the category “improving quality of care,” health care professionals now establish more contacts with other professionals working with patients with COPD at their health care center. Furthermore, considering the complexity of COPD management and the importance of an interdisciplinary treatment approach [2,55], if eHealth tools can facilitate interprofessional collaborations, eHealth could be used to tackle a key obstacle (low access to evidence-based care) to improving the quality of COPD services and care [56,57]. eHealth tools such as the COPD Web could be an essential means of reducing hierarchies, skepticism, and lack of knowledge about how other professions work and their roles in COPD management, both of which are known barriers to collegial and interprofessional collaboration [58,59].

Lastly, when implementing and using an eHealth tool for health care professionals as part of their clinical practice, our results align with previous research suggesting that there are likely to be initial barriers requiring time and education that need to be addressed [16,60-62]. For example, the category “efforts required for implementation” expressed a need for support and guidance, especially because additional time is necessary for learning to use the eHealth tool. Similar to our findings, a systematic review by Koivunen and Saranto [60] found that inadequate support and lack of training were critical obstacles to implementing and using eHealth tools among health care professionals. Comparable findings were expressed in a recent Cochrane qualitative evidence synthesis [61] on mobile eHealth tools, highlighting health care professionals’ need for education, support, and training when considering eHealth tools within clinical practice [60,61]. The need for initial training and education has also been reported among health care professionals involved in COPD management [62-64]; for example, Brewster et al [64] found that practical training and education were repeatedly seen as a facilitator for health care professionals to accepting and using eHealth tools and technologies.

Notably, the health care professionals in this study expressed that the biggest obstacle to using the eHealth tool was unrelated to the health care professionals themselves. Instead, it was related to whether patients did or did not have a computer at home or whether their computer skills were (perceived to be) too low. Several studies in the systematic review authored by Koivunen and Saranto [60] and other studies of various chronic diseases, including COPD [65-67], have highlighted that lack of access and skills are obstacles to using eHealth tools among patients. For example, we recently found that among patients with COPD enrolled in primary care, about 40% of eligible patients with COPD declined participation in an eHealth intervention due to no/limited experience with computers [17]. Furthermore, even among those accepting participation, and thus likely to consider themselves armed with sufficient technological skills, it was found that a higher need for technical support was identified as a primary barrier to usage among nonusers/seldom users of the eHealth tool during a 3-month intervention period [18]. Importantly, although it might be a primary barrier, we know from previous work on eHealth use among patients with COPD that continued use of eHealth tools among patients over time enables a transition from being insecure and experiencing technical concerns to acquiring technical confidence and improving disease management [68]. Nevertheless, it should be noted that the potential lack of access to computers and low computer skills among patients are not certain, as these were related via the experiences of health care professionals and were not expressed directly by patients. In a recent study, Sönnerfors et al [69] found that among patients with COPD in Sweden, over 90% had access to the internet, and 68% had access to a computer or laptop. Participants also had high knowledge of how to use the internet, with 91% having used the internet during the last 3 months and 85% almost every day. Taken together, indicating that although low access and computer skills are obstacles to using eHealth tools across various patient groups [65-67], it is vital that health care professionals do not draw firm conclusions based on their perception of the computer skills of their patients. Instead, a more relevant alternative for health care professionals would be to assess the level of health literacy among their patients to aid them in deciding whether the incorporation of an eHealth tool would be feasible. Health literacy has been identified as a vital explanatory factor affecting the usage of eHealth tools over time among patients with COPD.

Strengths and Limitations
Methodological strengths in this study are its design following the COREQ guidelines, increasing the credibility of our findings [16]; that our interviewees were multidisciplinary regarding health care professions; and that no extra resources were provided to the primary care units as health care professionals used the COPD Web as a part of their regular work practice [10]. Furthermore, throughout the analysis process, triangulation between researchers with different backgrounds was used to achieve higher credibility [32], and all authors were involved in creating subcategories and categories, and changes were made until consensus was achieved. In addition, several strategies have been used to enhance trustworthiness [32,33]. First, interviews were conducted via a face-to-face meeting at the health care professional(s) workplace at 3 months and over the telephone at 12 months due to practical choices. The 2 interview methods are considered equally credible [47]. Interview times vary and, occasionally, are short. Still, we interpreted our data to be rich enough for the analysis performed here, and a specific duration is not a guarantee for richness [46,48]. Second, during the analysis, triangulation between authors was made to ensure that our interpretation was grounded in the empirical data [46]. In addition, we continuously consulted the audio recordings when triangulation indicated risks of interpretational differences in a transcript [29]. Lastly, even though the number of interviews is not a crucial criterion in qualitative methods, it should be noted that the number of health care professionals, especially at 12-month interviews, was small. Although the 12-month interviews were fewer and shorter, they were included in the analysis following the study protocol [10]. Notably, the 12-month interviews did enrich the material, providing an
important insight that among those who continued to use the COPD Web, the eHealth tool seemed more integrated into daily work than it had been after the initial 3-month period, aligning with previous research highlighting health care professionals’ need for time, as well as education, support, and training when implementing eHealth tools in clinical practice [60,61]. By contrast, we also found that although the eHealth tool was mainly considered positive during the initial 3-month follow-up, 4 out of 10 health care professionals did not use the tool at 12 months, thus indicating that additional strategies might be necessary to implement the tool in clinical practice successfully.

Conclusions
This study is among the first to explore experiences of using an eHealth tool among health care professionals involved in COPD management. Our novel findings highlight that using an eHealth tool such as the COPD Web was experienced as providing knowledge support for health care professionals, leading to adaptation and facilitation of working procedures and person-centered care, enhanced patient contact, and encouragement of interprofessional collaboration—altogether improving quality of care. Furthermore, health care professionals emphasized that patients using the COPD Web were experienced to be better equipped to tackle their disease and adhere better to treatment—also increasing patients’ ability to self-manage their care. Lastly, before an eHealth tool can be successfully implemented within daily praxis, structural and external barriers requiring time, support, and education need to be addressed.

Data Availability
The data sets generated and analyzed during this study are not publicly available to protect the participants’ confidentiality. However, they are available from corresponding author upon reasonable request.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Guide to interviewing staff who have used the COPD Web. COPD: chronic obstructive pulmonary disease.

References


Removing Dust From the German Health Care System by Introducing Health Apps Into Standard Care: Semistructured Interview Study

Alexandra Heidel1; Christian Hagist1; Stefan Spinler1; Michael Schoeneberger1
WHU - Otto Beisheim School of Management, Vallendar, Germany

Corresponding Author:
Alexandra Heidel
WHU - Otto Beisheim School of Management
Burgplatz 2
Vallendar, 56179
Germany
Phone: 49 261 6509 25
Email: alexandra.heidel@whu.edu

Abstract

Background: In 2019, Germany launched the Digital Healthcare Act. The reform enables physicians to prescribe health apps as treatments to their statutory-insured patients.

Objective: We aimed to determine the extent to which the integration of health apps into standard care could be considered beneficial and which aspects of the regulation could still be improved.

Methods: We conducted a semistructured interview study with 23 stakeholders in Germany and analyzed them thematically. We used descriptive coding for the first-order codes and pattern coding for the second-order codes.

Results: We created 79 first-order codes and 9 second-order codes following the interview study. Most stakeholders argued that the option of prescribing health apps could improve treatment quality.

Conclusions: The inclusion of health apps into German standard care could improve the quality of treatment by expanding treatment portfolios. The educational elements of the apps might additionally lead to more patient emancipation through a better understanding of personal conditions. Location and time flexibility are the biggest advantages of the new technologies, but they also raise the most significant concerns for stakeholders because app use requires personal initiative and self-motivation. Overall, stakeholders agree that the Digital Healthcare Act has the potential to remove dust from the German health care system.

(JMIR Hum Factors 2023;10:e42186) doi:10.2196/42186

KEYWORDS
health apps; DVG; Digitale Versorgung Gesetz; Digital Healthcare Act; mobile health; mHealth; German statutory health care system; interview study

Introduction

Overview

In 2019, Germany was the first country worldwide to launch an act that enabled medical doctors to prescribe health apps as treatments to their patients—the so-called Digital Healthcare Act (Digitale Versorgung Gesetz [DVG]) [1]. Therefore, health apps became part of the German standard health benefit basket, financed by the statutory sickness funds. Previously, health apps were offered on a voluntary and discretionary basis in Germany, depending on the decisions of individual sickness funds or private health insurance companies. The statutory health care system in general was not covering the costs of any health app. Now, health apps can enter a preceding certification process by the “Federal Institute for Drugs and Medical Devices” (Bundesinstitut für Arzneimittel und Medizinprodukte [BfArM]). If the certification process is successful, the health app becomes a so-called “digital health app” (Digitale Gesundheitsanwendung [DiGA]).

The validation and certification process for these health apps is an entirely new process and still leaves room for future research and discussion [2]. The Digital Healthcare Act has the potential to decrease the costs associated with unnecessary doctor’s visits and substitute or complement other traditional treatments.
through digital initiatives related to patient education and self-management [3].

The BfArM has received 161 applications for admission to the DiGA index by January 2023, which would sanction these apps as prescribable treatments [4]. In January 2023, already 40 health apps were listed in the DiGA index, and they are now available via a physician’s prescription [5]. Many countries, especially in Europe, are observing the DiGA development in Germany closely, as they aim to introduce similar reimbursement strategies to disburden the health care system and increase the level of digitization of standard care. Belgium and France identified as following the German DiGA reimbursement example [6].

**Certification Process of DiGA**

The certification process for health apps and digital health devices was specified within the digital device regulation (Digitale-Gesundheitsanwendungen-Verordnung [DiGAV]) [7]. The BfArM published guidelines for health apps based on § 139e clause 8 (1) German social code (Sozialgesetzbuch [SGB V]) [8]. The guidelines highlight that DiGA need to be medical devices of the risk classes I or IIa, according to European Union regulation 2017/745 [9]. The guidelines explain the procedure for admission to the DiGA index. First, the app provider needs to apply to the BfArM to be admitted to the official index of reimbursable DiGA. The BfArM then examines the app or the digital health device for safety, quality, data security, data privacy, and several functional requirements within a 3-month period after the application was submitted. Thereafter, the BfArM conducts a first assessment of the potential positive treatment effects of the app. If this evidence is not yet sufficiently demonstrated in studies and publications but all other requirements are fulfilled, the health app may still receive preliminary acceptance to the index according to § 139e SGB V [2]. During this phase, the health app is in a 12-month test phase.

The app can be prescribed through medical doctors during the test phase, and the health app provider may set the price for market entry. After 12 months, the health app provider needs to demonstrate sufficient proof of positive care effects. The legislator used the term positive care effect in the DVG and defined the concept as a medical beneficial outcome or patient-relevant procedural improvement in care [1]. If sufficient proof of a positive care effect cannot be demonstrated, the app is removed from the index, and a prescription is no longer possible. If the health app provider has demonstrated sufficient effectiveness, the price for use of the app is negotiated with the national association of statutory health sickness funds [2]. This system of preliminary market access and reimbursement is supposed to facilitate innovation within the health care sector. After negotiating the final price, the app is permanently accepted to the DiGA index [2].

**The DiGA Prescription Process**

The DVG is one of many initiatives by the German Federal Ministry of Health to modernize and digitize the German health care system. The aim of the act is to quickly introduce innovative digital treatment solutions into the standard care portfolio and to give statutory sickness funds the opportunity to encourage more efficiency and higher quality treatment [1]. The DVG enabled statutory health–insured patients to claim digital solutions, if available, for disease management and treatment. Physicians, as the gatekeepers of the German health care system, play a major role in the success of the DVG. According to the act, physicians are required to recommend and prescribe suitable health apps and supervise the app use of the patients according to their individual disease progression [1]. Compensation for this supervision is not yet sufficiently regulated. Hence, the reform contains a subsection stating that practitioners’ efforts shall be compensated, but a clear guideline and incentive system is yet to be negotiated [1].

In May 2020, the board of the German Medical Association recommended compensation for practitioners prescribing and providing advice upon first-time use of a specific DiGA, according to the billing code for practitioners (Gebührenordnungsposten [GOP]) as GOP 01470 [10]. This code reimburses the practitioner an amount of 2.00 € (US $2.21) and may only be billed once per app [10]. Just recently, a new billing code numbered 86700 has been introduced to reimburse practitioners to monitor twice a year the progress of the app use. However, not all medical specialist groups, such as urologists, were included in the compensation logic; they are, therefore, not allowed to use the billing code for supervision [11]. This is a symbolic starting point but might not be enough to set an effective incentive system for practitioners.

**The German Ambulatory Setting**

In Germany, most physicians in the ambulatory sector are self-employed. Their motivation to enhance and recommend the use of health apps might also be debatable given the lack of financial incentives to do so. Many private practitioners lack a range of digital solutions in their practices [12]. Approximately, only 56% to 58% of German private practitioners have already digitized processes, such as patient documentation, appointment planning, and waiting time management, for their practices [12]. Just 37% of resident doctors are willing to standardize their patient documentation to accelerate the introduction of a digital patient file to encourage better patient data exchange between different specialties [12]. In Germany, there is an imbalance between the demand and supply of physicians, partly explained by a general shortage of physicians, especially in rural areas, and partly explained by the unique statutory health care system and the apparent nearly unlimited and free doctor’s treatment portfolio for statutory health–insured patients [13].

Conducting an interview study, we aimed to determine the extent to which the integration of health apps into standard care could be considered beneficial by different stakeholders and which aspects of the regulation could still be improved. We expected a general reticence toward the DVG from most of the stakeholders. Yet we also expected that the acceptance of app treatments is currently changing due to the experiences of the COVID-19 crisis since location-independent, flexible, and at-home practicable solutions have gained importance.
Methods

Procedure
We used an interview study approach to explore different aspects of the introduction of mobile health services in the German statutory health care system. We conducted a semistructured interview study with 23 stakeholders in Germany and thematically analyzed those interviews [14].

First, we identified relevant stakeholder groups to guide sampling. The stakeholder groups are the following:

- Certification institutions: institution that currently and in the past examined and certified medical devices, digital preventive care solutions, or DiGA.
- Medical doctors: physicians who work in the ambulatory sector in different specialties.
- Health app producers: companies that develop digital medical solutions.
- Statutory sickness funds representatives: representatives who work for statutory sickness funds within a DiGA business unit or project group.
- Political representatives: politicians who work for regional or federal ministries.
- Medical chamber representatives: representatives who work for different regional medical chambers, which are compulsory institutions that represent the interests of physicians in Germany.

We contacted 65 stakeholders via purposeful sampling based on their profession and expertise between October 2019 and December 2019 [15]. Thereafter, 23/65 stakeholders responded to and participated in the study. Second, we created a suitable interview guide (Multimedia Appendix 1), discussed, and tested the questions in a real interview scenario with a previously selected stakeholder. We conducted the interviews between October 2019 and January 2020 with a certification body representative (1/23), medical doctors (9/23), health app producers (2/23), medical chambers (4/23), political representatives (5/23), and statutory sickness funds’ representatives (2/23). Most interview partners were middle-aged (Table 1) and almost equally distributed by gender (13/23 were male and 10/23 were female).

We used the software ATLAS.ti (version 9.0.18; ATLAS.ti Scientific Software Development GmbH) to thematically analyze and cluster the transcripts. Two researchers independently coded the transcripts using thematically relevant first- and second-order codes and found consensus about the final codes by merging the coding data, and therefore, consolidating the most important themes (final codes) through educated discussions [16]. The procedure to establish first-order codes consisted of highlighting the important parts of the transcripts and summarizing these through descriptive first-order codes [16-19]. In the second step, we aggregated the descriptive first-order codes so that all duplicates could be removed without any loss of important information. In the final step, we used pattern coding to organize and cluster the second-order codes by the most relevant topics [16,20].

Table 1. Descriptive statistics interview study.

<table>
<thead>
<tr>
<th>Descriptive statistics</th>
<th>Gender, n/N (%)</th>
<th>Age group (years), n/N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Female</td>
<td>Male</td>
</tr>
<tr>
<td>Medical doctors</td>
<td>4/9 (53)</td>
<td>5/9 (47)</td>
</tr>
<tr>
<td>Statutory sickness funds’ representatives</td>
<td>1/2 (50)</td>
<td>1/2 (50)</td>
</tr>
<tr>
<td>App certification representative</td>
<td>1/1 (100)</td>
<td>0/1 (0)</td>
</tr>
<tr>
<td>Medical chamber representatives</td>
<td>0/4 (0)</td>
<td>4/4 (100)</td>
</tr>
<tr>
<td>Political representatives</td>
<td>2/5 (40)</td>
<td>3/5 (60)</td>
</tr>
<tr>
<td>Health app producers</td>
<td>2/2 (100)</td>
<td>0/2 (0)</td>
</tr>
</tbody>
</table>

Ethical Considerations
The ethics approval is not applicable to this study, as we conducted expert interviews. Participants consented the content of the questions. We followed the ESOMAR international code on marketing, opinion, social research, and data analytics [21]. During the expert interviews, we did not ask any personal or confidential content. All questions were subject to health care professional content. All stakeholders agreed in the beginning of the interview to the collection of data and were informed that the pseudonymized transcripts of the interviews are going to be stored at our university server in Germany. No sensitive or personal data were collected.

Results
Quantitative Results
Within the first coding round, we identified 1048 first-order codes. After discussing their meaning, we merged these codes into 79 first-order codes. Finally, the first-order codes were clustered into 9 second-order codes. These are depicted in Table 2. The interviewee overview and the ATLAS.ti code report, depicting all first- and second-order codes, can be found in the Multimedia Appendix 2.
Table 2. Second-order codes. Our calculation was based on ATLAS.ti coding protocol.

<table>
<thead>
<tr>
<th>Second-order code</th>
<th>Frequency first-order codes (n=79), n</th>
<th>Total number of quotes, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor patient and potential care effects</td>
<td>12</td>
<td>200</td>
</tr>
<tr>
<td>Certification process</td>
<td>9</td>
<td>112</td>
</tr>
<tr>
<td>Chances for the health care system</td>
<td>7</td>
<td>93</td>
</tr>
<tr>
<td>Cost development</td>
<td>7</td>
<td>61</td>
</tr>
<tr>
<td>Factor doctor and potential effects on daily routine</td>
<td>14</td>
<td>293</td>
</tr>
<tr>
<td>Political incentive systems</td>
<td>8</td>
<td>78</td>
</tr>
<tr>
<td>Role of the statutory health insurer and reimbursement</td>
<td>7</td>
<td>58</td>
</tr>
<tr>
<td>Considerations for the app developers</td>
<td>5</td>
<td>28</td>
</tr>
<tr>
<td>Concerns about data use, data privacy, and data security</td>
<td>10</td>
<td>108</td>
</tr>
</tbody>
</table>

Chances for the Health Care System

We observed a generally positive perception of the DVG and the option of prescribing health apps as treatments during the interview study. However, most stakeholders would not want to overestimate the effect of health apps introduced as treatments in the German health care system. Medical doctors thought that prescribed health apps should be regarded as optional treatments and not as replacements or substitutions for traditional treatments.

Factor Patient and Potential Care Effects

The majority of stakeholders thought that the additional option of prescribing health apps could improve treatment quality for patients. The use of health apps has various positive effects for patients, such as more flexibility in terms of location and time as well as a permanent reduction in waiting time for appointments. Most stakeholders argued that the use of health apps could lead to patient emancipation through better disease education and management. It was said that “especially chronic patients could benefit if they need permanent guidance.” However, medical doctors were especially concerned that patients might not use or might incorrectly use the app-based treatment. Therefore, app use supervision and advice from medical doctors should be indispensable. One of the respondents said, “It is important that these technologies are just used with medical supervision, especially for risk patients.” Another concern was that many patients could be excluded from the app treatments because of demographic factors, such as age or local internet connection. One of the respondents said, “An elderly woman aged 70 years—I do not know if she would use these technologies.”

Factor Doctor and Potential Effects on Daily Routine

Many stakeholders argued that the prescription of DiGA could enhance the service portfolio of resident doctors. One of the respondents said: “I think that a quality improvement of care is a possible outcome”. Many medical doctors would be delighted if health app use would lead to fewer unnecessary doctor’s visits and therefore again increase treatment time for patients with severe or complicated conditions. Furthermore, medical doctors would have the chance to detect chronic or severe conditions earlier through data insights, which they would not be able to obtain from traditional treatments or patient disease management systems.

Certification Process

However, the unique certification process might not only be a chance for improvement and innovation but also an opportunity to abuse the system by very high price settings. This could lead to short-term cost increases within the German healthcare system.

Costs Development

To prevent expensive app collection without use from patients, some stakeholders suggested monitoring compliance and letting patients pay for prescribed apps if they do not use them. On the other hand, statutory-financed health apps also foster the use and perception of health apps in general within society. Technologies such as gamification and nudging may increase patient compliance and use even further for specific treatments.

Considerations for the App Developers

Some stakeholders recommended a pay-for-performance principle, which means that the final costs for the app should depend on the intensity of the real positive care effect verified during the one-year test phase.

Political Incentive Systems

There is no sufficiently regulated incentive or remuneration system for physicians who would have an increased workload because of continuous app supervision. Stakeholders from all sectors of the health care system recommended the introduction of individual billing codes and an appealing remuneration system for physicians who supervise app treatments because they fear a blockage of the innovation.

Concerns About Data Use, Data Privacy, and Data Security

Many stakeholders fear a lack of data security and data privacy for patients; medical doctors especially question the responsibility in cases of data theft and severe personal consequences for patients.
Role of the Statutory Health Insurer and Reimbursement

All stakeholders recommended that statutory sickness funds, health app producers, and medical chambers in particular should offer a wide portfolio of health app education initiatives to address the needs and interests of physicians with different specialties, ages, location characteristics, and different patient clientele. Therefore, one of the respondents demanded “more education, even workshops about digital treatment solutions because this is important.”

Figure 1 presents the main findings and recommendations from the interview study, embedded in the regulatory framework of DiGA certification and implementation.

Figure 1. Benefits, risks and recommendations for the Digital Healthcare Act (Digitale Versorgung Gesetz [DVG]). DiGA: Digitale Gesundheitsanwendung (digital health app); MDR: Medical Device Regulation.

Discussion

Mainly, we identified a relative openness toward the introduction of DiGA into standard care. Yet there have been some concerns as well, regarding data security, compensation of medical doctors, and the self-motivation of patients. However, most stakeholders expected benefits resulting from the introduction of the DVG.

One of the major concerns identified during the interview study was that health apps might not provide the desired positive care effects, and therefore, could lead to an unnecessary short-term increase in costs for the German health care system. However, compliance is not just an inhibitor to improvement in the digital sphere but also in the analog treatment world. In particular, medical doctors expressed their concerns in the interview study that digital treatments could lead to a short-term cost increase because the app treatments require self-motivation, which has also been argued by Safi et al [22]. Yet many studies disagree with this standpoint because modern technologies, such as gamification and nudging, have shown a significant positive effect on patient compliance [23,24].

A major advantage of the app treatment versus the traditional treatment is that patients gain location and time flexibility. According to most stakeholders, this advancement could lead to an improvement in treatment quality and service due to an extension of health care portfolios. Dahlhausen et al [25] came to similar conclusions resulting from their survey about DiGA with German practitioners.

All stakeholders agreed that there is a need to introduce an appealing and individual financial incentive system to remunerate the increased workload that medical practitioners have due to continuous app advice, supervision, and data analysis. All stakeholders proposed individual billing codes for practitioners based on workload increase to ensure the support and participation of these important gatekeepers.

The opportunities that app treatments offer through data generation and patient monitoring could improve research and diagnostics to a large extent because of their regular real-world and real behavioral documentation [26]. App treatments are not supposed to replace traditional treatments, but app-based treatments offer many opportunities and additional benefits, which is why app-based treatments should be regarded as a valuable complement to medical care portfolios [25]. Yet a representative survey with practitioners showed that 33.6% of the participating physicians have already prescribed a DiGA in 2022 [27]. In 2021, just 14.3% prescribed DiGA; and in 2020, just 1% did so [27]. This means we see a fast adoption rate, and contrary to our hypothesis, a general openness to prescribe and use DiGA in the standard care setting.
The educational elements of the apps might additionally lead to more patient emancipation through a better understanding of personal conditions. Location and time flexibility are the biggest advantages of the new technologies, but they also raise the most significant concerns for stakeholders because app use requires personal initiative and self-motivation, as also argued by Weise et al [28]. Physicians should supervise and monitor patients' app use to support the adequate use of the app as a treatment. This supervision might also help to prevent patients from collecting but not using reimbursable health apps, and therefore, exploiting the system.

We conducted this study prior to the COVID-19 crisis. Hence, we now expect an increased positive perception of DiGA due to experiences during the lockdown in Germany introduced on March 25, 2020.

In conclusion, stakeholders within the German health care system had generally an open mind toward the Digital Healthcare Act and felt that the introduction of the act helps to relieve the dust from the German health care system and pushes forward the digitization of the industry. The inclusion of health apps in the statutory health care system could improve the quality and service of treatment by expanding portfolios, and therefore, is considered beneficial by most stakeholders.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Interview questionnaire guide.
[DOCX File, 16 KB - humanfactors_v10i1e42186_app1.docx ]

Multimedia Appendix 2
Interview partner overview and code report.
[DOCX File, 21 KB - humanfactors_v10i1e42186_app2.docx ]

References
8. Sozialgesetzbuch (SGB) Fünftes Buch (V). § 139e Verzeichnis für digitale Gesundheitsanwendungen; Verordnungsermächtigung.: Bundesamt für Justiz URL: [https://www.gesetze-im-internet.de/sgb_v/§_139e.html](https://www.gesetze-im-internet.de/sgb_v/§_139e.html) [accessed 2021-10-12]
12. Albrecht M. Praxisbarometer Digitalisierung: IGES Institut in Zusammenarbeit mit otten software. IGES Institüt.: KBV; 2018. URL: [https://www.kbv.de/media/sp/PraxisBarometer-Digitalisierung_IGES-Praesentation.pdf](https://www.kbv.de/media/sp/PraxisBarometer-Digitalisierung_IGES-Praesentation.pdf) [accessed 2023-04-20]


Abbreviations

- BfArM: Bundesinstitut für Arzneimittel und Medizinprodukte (Federal Institute for Drugs and Medical Devices)
- DiGA: Digitale Gesundheitsanwendung (digital health app)
- DiGAV: Digitale-Gesundheitsanwendungen-Verordnung (digital device regulation)
- DVG: Digitale Versorgung Gesetz (Digital Healthcare Act)
- GOP: Gebührenordnungsposition (billing code)
- SGB: Sozialgesetzbuch (German social code)
Optometrists' Perspectives Regarding Artificial Intelligence Aids and Contributing Retinal Images to a Repository: Web-Based Interview Study

Aurora Constantin¹, PhD; Malcolm Atkinson¹, PhD; Miguel Oscar Bernabeu², MSc, DPhil; Fiona Buckmaster³, MSc; Baljean Dhillon³, BM, BS; Alice McTrusty³, BSc, PhD; Niall Strang⁴, BSc, PhD; Robin Williams⁵, PhD

¹School of Informatics, University of Edinburgh, Edinburgh, United Kingdom
²Centre for Medical Informatics, Usher Institute, University of Edinburgh, Edinburgh, United Kingdom
³Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, United Kingdom
⁴Department of Vision Sciences, Glasgow Caledonian University, Glasgow, United Kingdom
⁵School of Social and Political Science, University of Edinburgh, Edinburgh, United Kingdom

Corresponding Author:
Niall Strang, BSc, PhD
Department of Vision Sciences
Glasgow Caledonian University
Cowcaddens Road
Glasgow, G40BA
United Kingdom
Phone: 44 07794835467
Email: n.strang@gcu.ac.uk

Abstract

Background: A repository of retinal images for research is being established in Scotland. It will permit researchers to validate, tune, and refine artificial intelligence (AI) decision-support algorithms to accelerate safe deployment in Scottish optometry and beyond. Research demonstrates the potential of AI systems in optometry and ophthalmology, though they are not yet widely adopted.

Objective: In this study, 18 optometrists were interviewed to (1) identify their expectations and concerns about the national image research repository and their use of AI decision support and (2) gather their suggestions for improving eye health care. The goal was to clarify attitudes among optometrists delivering primary eye care with respect to contributing their patients’ images and to using AI assistance. These attitudes are less well studied in primary care contexts. Five ophthalmologists were interviewed to discover their interactions with optometrists.

Methods: Between March and August 2021, 23 semistructured interviews were conducted online lasting for 30-60 minutes. Transcribed and pseudonymized recordings were analyzed using thematic analysis.

Results: All optometrists supported contributing retinal images to form an extensive and long-running research repository. Our main findings are summarized as follows. Optometrists were willing to share images of their patients’ eyes but expressed concern about technical difficulties, lack of standardization, and the effort involved. Those interviewed thought that sharing digital images would improve collaboration between optometrists and ophthalmologists, for example, during referral to secondary health care. Optometrists welcomed an expanded primary care role in diagnosis and management of diseases by exploiting new technologies and anticipated significant health benefits. Optometrists welcomed AI assistance but insisted that it should not reduce their role and responsibilities.

Conclusions: Our investigation focusing on optometrists is novel because most similar studies on AI assistance were performed in hospital settings. Our findings are consistent with those of studies with professionals in ophthalmology and other medical disciplines: showing near universal willingness to use AI to improve health care, alongside concerns over training, costs, responsibilities, skill retention, data sharing, and disruptions to professional practices. Our study on optometrists’ willingness to contribute images to a research repository introduces a new aspect; they hope that a digital image sharing infrastructure will facilitate service integration.
**KEYWORDS**

AI in optometry; repository of ocular images; user studies; ophthalmologists; AI; research; medical; decision support; tool; AI decision support tools; perspectives of optometrists and ophthalmologists; perspectives of optometrists and ophthalmologists; digital tool; digital decision support tool.

**Introduction**

Community optometrists in Scotland are being asked to contribute their collections of retinal images to a National Health Service (NHS) repository to enable research focusing on the earlier stages of eye diseases. This should enable improvements in the detection and treatment of those conditions. Optometrists are the first port of call for people with an eye problem as “optometrists (as graduates) are trained to examine the eyes to detect defects in vision, signs of injury, ocular diseases or abnormality and problems with general health, such as high blood pressure or diabetes. They make a health assessment, offer clinical advice, prescribe spectacles or contact lenses, and refer patients for further treatment, when necessary” [1]. We asked optometrists for their thoughts about contributing their patients’ data and their expectations about potential benefits and challenges.

Clinical research on diagnosis and treatment of eye diseases is largely confined to hospital ophthalmology services and universities. This is problematic; for example, clinical trial recruitment fails to reach individuals whose eye conditions (e.g., dry age-related macular degeneration) fall outside a referable disease threshold. To overcome this barrier, the Scottish Collaborative Optometry-Ophthalmology Network e-research (SCONe) seeks to create a repository of retinal images captured by optometrists in the community [2]. This will facilitate new clinical research. SCONe will also be an educational resource for auditing false-positive and false-negative referrals and provide exemplars, variants, and outliers. The former will improve patients’ pathways between primary and secondary care. The latter will improve clinical image interpretation. SCONe’s image repository will enable carefully governed research spanning the full diversity of Scottish patients and all stages of disorders. Images will be gathered from optometry practices, with 2 substantial benefits: (1) a nearly complete coverage of the population attending primary care optometry services and (2) coverage of a broad spectrum of disease severities, including early and undiagnosed disease as well as those with no disease. Information about diagnosis, treatments, and outcomes will depend on pseudonymized linkage to standard health care data sets.

The advent of the SCONe repository and the need to build back better (a UK rallying cry after COVID-19) motivated our study. We focused on optometrists because they are in the front line of eye health care, and in Scotland, they are provided extra training and NHS-provided cameras. Nearly 900 optometry practices employ 1300 optometrists in Scotland (Table 1). Optometrists refer 3%-9% of their patients to hospitals, with 89%-97% accuracy in Scotland [3]. Their wide distribution makes them more accessible to patients than hospital services. Optometrists’ collaboration with ophthalmologists is crucial. Therefore, we interviewed a small sample of ophthalmologists as well as optometrists to better understand their working relationship with optometrists.

<table>
<thead>
<tr>
<th>Practice type</th>
<th>Practices (n)</th>
<th>Description of practice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domiciliary</td>
<td>76</td>
<td>business providing eye care in a patient’s home or care setting</td>
</tr>
<tr>
<td>Independent</td>
<td>473</td>
<td>an individual or small group of locally owned optometry practices</td>
</tr>
<tr>
<td>Multiple</td>
<td>299</td>
<td>part of a large (typically national) chain or franchise of practices, for example, Specsavers or Boots Opticians</td>
</tr>
</tbody>
</table>

SCONe will improve artificial intelligence (AI) methods by exploiting the contributed images. The unprecedented population diversity and coverage will permit data-driven training and validation that potentially addresses recently highlighted issues of bias in AI methods. The images represent patients’ histories preceding recognized onset of eye conditions. They may therefore contain latent information that would have enabled earlier diagnoses. Such early predictors would be significant in several of the eye conditions identified by Campbell et al [5], where AI has the potential to improve eye care. The validation against the full population addresses one impediment to the uptake of AI-enabled methods identified by González-Gonzalo et al [6]. They propose a multi-step strategy, whereby carefully chosen sets of relevant stakeholders are fully engaged through all 7 stages—from planning to operation. Both these papers draw attention to the difficulty of moving from research to practical widespread use. Our interviewees’ perceptions of that challenge are a significant element of González-Gonzalo et al’s [6] first stage.

Given the aging population and the growing numbers and range of conditions that can now be treated, optometrists, as the eye-care front line, need help—potentially from AI. Our study focuses on optometrists to better understand their needs and constraints as they consider contributing patients’ images to a shared repository and prepare to use AI-powered assistance in their expanding primary eye care role. Although optometry services in Scotland have some particularities [7], our findings can be generalized to most countries.
Methods

Overview

This study employs semi-structured interviews to identify the attitudes to changes in the use of retinal images among practicing optometrists and ophthalmologists, particularly with respect to changes stimulated by SCONe. We conducted 23 web-based interviews between March and August 2021 via Microsoft Teams. Each interview lasted for 30-60 minutes. Only the interviewee and 1 interviewer (the first author) were present, except for the first ophthalmologist, while 2 other authors attended to refining the script for ophthalmologists.

The interview script was designed to reveal the expectations, issues, and constraints encountered by optometrists, for example, their uses of the shared images and worries over extra work and training that contributing images might require. The interview scripts were revised in consultation with our expert advisors and after 2 pilot interviews retaining topical consistency, as per a previous study report [8]. The questions covered 4 categories: (1) image sharing, (2) AI-enabled methods, (3) research, and (4) education/training related to image sharing and AI.

We also conducted a limited set of interviews with ophthalmologists to explore the crucial optometry-ophthalmology collaboration—primarily during referral but also during training and when negotiating revisions of responsibilities. The script for ophthalmologists was revised drawing on experience from 8 optometrist interviews retaining their topics and adding 2 new topics: (5) opening questions and (6) ophthalmology-optometry relationships (as per a previous study) [8].

We interviewed 23 people (18 optometrists of which 9 were females and 5 ophthalmologists of which 4 were females). Initially, we recruited interviewees through SCONe and extended and diversified our sample through snowball techniques and other channels. We covered a representative sample of optometrists that included both smaller independent practices and larger multiple practices (eg, national chains). The interviews included 14 optometrists from independent practices and 4 optometrists from multiple practices, with coverage of the diversity of practice contexts shown in Table 2. Recruitment was slow because interviewees had excessive workloads handling the pandemic backlog, which, given our completion deadline, restricted the number and range of recruits. However, toward the end, interviews revealed very few new issues; so, we believe we have validated our scripts and methods and can provide a good representation of optometrists’ views. For the ophthalmologists, we tried to cover various specialisms, for example, age-related macular degeneration, diabetic retinopathy, and glaucoma.

Table 2. Optometry interview coverage of Scottish urban/rural categories [9] showing ratios of interviewees in each category compared with the population sizes and number of optometry practices

<table>
<thead>
<tr>
<th>Category</th>
<th>Optometrists (n=18), n (%)</th>
<th>Population size (n=5,454,000), n (%)</th>
<th>Optometry practices (n=900), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large urban area</td>
<td>12 (66.6)</td>
<td>1,887,000 (34.6)</td>
<td>337 (37.5)</td>
</tr>
<tr>
<td>Other urban area</td>
<td>3 (16.6)</td>
<td>1,974,000 (36.2)</td>
<td>394 (43.8)</td>
</tr>
<tr>
<td>Accessible small town</td>
<td>1 (5.6)</td>
<td>464,000 (8.5)</td>
<td>72 (8)</td>
</tr>
<tr>
<td>Remote small town</td>
<td>1 (5.6)</td>
<td>191,000 (3.5)</td>
<td>60 (6.7)</td>
</tr>
<tr>
<td>Accessible rural area</td>
<td>1 (5.6)</td>
<td>611,000 (11.2)</td>
<td>13 (1.5)</td>
</tr>
<tr>
<td>Remote rural area</td>
<td>0 (0)</td>
<td>322,000 (5.9)</td>
<td>24 (2.6)</td>
</tr>
</tbody>
</table>

The interviews (total of 15 hours) were recorded and professionally transcribed, introducing pseudonyms to protect identities but permit follow-up studies. The analysis followed the thematic analysis method [10] using NVivo (Lumivero) [11]. Thematic analysis is one of the most widely used methods in qualitative studies. Its purpose is to identify themes (ie, patterns) in the data and relationships between themes that are relevant to a specific research topic/phenomenon. Thematic analysis is well-suited for analyzing large texts such as transcripts of a set of interviews. The next section presents the analysis of our data by using new methods and technologies notwithstanding worries about costs and workloads. Five themes came out from the analysis. Their order results from ordering topics in the scripts are as follows: (1) changes to professional working patterns, (2) envisaging the image repository’s impact, (3) benefits from AI decision support, (4) paths to improved eye care, and (5) education and training. We present our analysis structured by these themes summarizing significant views expressed, with the number of interviewees who supported each point, out of 18 for optometrists and 5 for ophthalmologists.

Ethics Approval

This study has been approved by the School of Informatics (University of Edinburgh) Research Ethics Committee (RT 62378).

Results

Overview

The interviews and analysis revealed consistent commitment to maintaining high professional standards and improving eye care by using new methods and technologies notwithstanding worries about costs and workloads. Five themes came out from the analysis. Their order results from ordering topics in the scripts are as follows: (1) changes to professional working patterns, (2) envisaging the image repository’s impact, (3) benefits from AI decision support, (4) paths to improved eye care, and (5) education and training. We present our analysis structured by these themes summarizing significant views expressed, with the number of interviewees who supported each point, out of 18 for optometrists and 5 for ophthalmologists.

Table 2. Optometry interview coverage of Scottish urban/rural categories [9] showing ratios of interviewees in each category compared with the population sizes and number of optometry practices
We present extracts of the transcripts retaining the abbreviations they used.

**Theme 1: Changes to Professional Working Patterns**

Funding pressures were a predominant issue as optometrists considered increasing their clinical responsibilities. Optometrists and ophthalmologists drew attention to the commercial pressures on optometrists in their competitive market of small private enterprises and larger multiples where prescribing spectacle frames and lenses is financially as well as clinically essential.

(12 optometrists, 3 ophthalmologists).

...I think sometimes optometry is strange because it is...90% clinical but there is a commercial aspect to it...competition with each other. [P3]

Those from independent practices felt they had more choice over their allocation of effort.

...Independent practices work very differently ...I don’t feel time pressured, ...if I need more time, I take more time .... the majority of optometrists, especially [those in] multiples, do not [have that luxury]. [P8]

The divergence between prescribing lenses and diagnosing other eye conditions may be reinforced by patient attitudes.

...people who don’t believe that there’s a problem with their eyes just want to go and get glasses. [P28]

Five optometrists expressed concerns over meeting the costs.

...the way the GOS [General Ophthalmic Services] contract is structured clinical stuff isn’t the thing that pays the bills. So, something that’s going to generate workload, but not potentially generate [income] is going to be a tough sell. [P2]

These choices are affected by public policy and funding. P24 explained that “The contract ... in Scotland empowers us more, pays us more, pays us to [monitor] conditions.” P25 noted that although they are doing more tests and interpretation of the results and spending time explaining these to patients, “the NHS fee hasn’t changed very much at all.” The initial fee reflected the cost, but it has not been increased in line with inflation and additional procedures. Optometrists in Scotland take more responsibilities in health care.

...I would say that that is less common in Scotland now [to have optometrists who do not want to do more than prescribe lenses] because we’ve been doing this sort of work ...for a long time now. ...There’s a budget for training and developing optometrists. [P24]

To reduce the burden on secondary care and to obtain good quality images, one of the ophthalmologists proposed the establishment of specialist imaging hubs.

...It would be very useful to have imaging hubs where imaging equipment can be standardized and similar to those used in NHS Ophthalmology Departments...high resolution photographs which are essential for safe management of patients. The resolution we want is virtually impossible for all the opticometric practices to have. [P22, ophthalmologist]

Eleven optometrists and 4 ophthalmologists proposed to expand the role of optometrists.

...Hopefully, it will enable us to provide a better service for patients. There might be times where you do not have to refer to hospitals and you will manage someone locally. [P15]

One optometrist believes “a huge number of optometrists are willing to take on more responsibility” but suggested that this could stimulate stratification of the profession.

...We need to start having some kind of differentiation in the hierarchy in eye care. Not just optometrists and ophthalmologists but a continuum between the two. [P28]

This expansion of roles should be carefully analyzed (as indicated by 4 optometrists and 2 ophthalmologists). For example, extra skills will be needed. However, optometrists may not receive additional remuneration.

...The idea behind it is that we’re moving more professionally and we’re moving into a better, more rewarding profession, but we’re not having enough money. [P25]

**Theme 2: Envisaging the Image Repository’s Impact**

All optometrists were keen to see an extensive and long-running research repository containing their patients’ images. Its primary role is to improve AI. However, its educational role and coordination may facilitate communication and collaboration. The following issues dominated: (1) professional relationships, (2) teleconsultations, (3) health inequalities, (4) image quality, (5) standardization and automation, (6) scrutiny out of context, and (7) the need for electronic health records (EHRs).

Interviewees expanded our vision of what mattered, which was a goal of our in-depth interviews. They highlighted the critical needs.

Most optometrists (12/18, 67%) hoped the image repository would facilitate supporting each other.

...We can learn from other people’s treatment management plans...we can enhance a collective learning and collective management of patients. [P18]

Optometrists (10/18, 56%) hoped the shared repository would stimulate a change in relationships between optometrists and ophthalmologists.

...A project like this could...make optometrists feel more part of the whole...linking primary and acute services so that you’re out in the community, but we are part of the hospital project and yes, we engage with this data gathering which suddenly gives you the message...your work is appreciated and valued. [P7]

Two optometrists hoped that it will mean hospital staff gain increased respect for optometry.

...The communication [between the optometrists and ophthalmologists] would necessarily have to increase ...hospital eye care needs to understand the importance of the role of the community care ...making their life easier. [P28]
This contrasts with ophthalmologists’ conviction that their “relationship with optometrists is very good” [P19, ophthalmologist]. Ophthalmologists were aware of plans to deploy an EHR system—a better pathway to using images when coordinating patient care.

...We’re supposed to be going to a digital, an electronic patient record with images, so if that was shared with optometrists, they might be able to learn from that. [P23, ophthalmologist]

Four local [12] optometrists and all the ophthalmologists in this study proposed using teleconsultation, possibly stimulated by its use during the COVID-19 pandemic [13].

...I could see it being really useful for...to bring eye health screening to more remote locations. [P28]

...But the benefits are at least timeously we can be getting in touch with the patient and either having a chat with them or reviewing images just to make sure that there is no gross alteration. [P22, ophthalmologist]

Teleconsultations should result in more accurate triage.

...I think the benefits of it would definitely be reducing the number of patients who need to be seen in the hospital...bring the images to the doctor instead of the whole patient. [P23, ophthalmologist]

One ophthalmologist pointed out the risks of virtual examinations.

...The risk, which I always explain to patients, in a virtual appointment is that it is easy to miss subtle changes in retinal pathology. [P22, ophthalmologist]

Three optometrists suggested that data sharing would reduce variations in the care.

...It might be useful, because in some areas there might be more healthy patients, and in others there might be unhollower patients. The retinal images might help get...understanding...managed more. [P4]

Five optometrists were concerned about the quality of the retinal images they produced. Causes included device quality, opacities (eg, cataract), and no dilation. In hospital, dilation is routine. In optometry practices, dilation is the only standard for patients aged over 60 years.

...And I think there’s also a difference, and I don’t know if this has really been considered, between imaging captured in hospital and imaging captured in practice, in that a lot of imaging captured in practice isn’t dilated, isn’t of the quality that is captured in a hospital or a screening service setting. [P2]

Three optometrists anticipated that the image repository would raise image quality concerns.

...I think just...standardizing things across the board so the people who are taking the images know. [P10]

Six optometrists and 3 ophthalmologists mentioned standardization for images and software applications that “talk to each other.”

Optometrists anticipated service improvements from image sharing. Two optometrists hoped image sharing would avoid redundant work and improve patient care.

...There’s far too much duplication of services...if we have a centralized service, and all this data is collected, it can only improve for patients...and the health of their eyes. [P2]

Four optometrists hoped for more accurate information by accessing patients’ records.

...So...rather than depending on a patient’s word of mouth...patients are not the best in relaying accurate information about their past treatments...it would be nice to be able to access the actual data. [P18]

Three optometrists expected the repository would help them track patients’ data.

...Patients move and patients’ care providers move, and patients often attend optometry where they work as opposed to where they live. [P2]

Some optometrists feared that an image repository would be used to investigate whether their decisions were right, “putting themselves at undue risk for litigation whenever their records are being pulled apart by other people” [P4].

Another optometrist worried about clinical decisions being scrutinized.

...So, I would be a little bit concerned that...someone would be looking over my shoulder and deciding whether I had made the right decision or not, whereas they didn’t have all the data that I had available. [P28]

Eight optometrists worried about the technical requirements and time needed to contribute images.

...That’s the biggest challenge, it’s always...how do we get these images...transferred easily, and that’s not too time consuming. [P5]

One optometrist wondered whether the contribution process could be automated.

**Theme 3: Benefits From AI Decision Support**

Considering the use of AI decision support, all optometrists anticipated benefits and were happy to use it, provided they retained the ultimate responsibility. They expected more treatment options with better guidance and help when they encounter something new. Their issues included (1) diagnostic skill acquisition and retention, (2) divergence in the optometry community, (3) changing relationships with patients, and (4) impatience over the rate of deployment. The AI tools were seen as augmenting their skills, empowering them to make better decisions (as stated by 14 optometrists).
Ten optometrists felt strongly that they should retain control and that clinical decisions must remain their responsibility. 

You have to take the human factors into account and base it on all your observations not just what the technology is telling you. [P15]

Ophthalmologists considered that AI decision support could help optometrists make better referrals and detect eye diseases earlier.

I have no doubt...[that AI tools help optometrists make better referrals]. It will help a lot...in early diagnosis...one of our main cornerstones in management of glaucoma. [P27, ophthalmologist]

More efficient interpretation of the growing number of images may be delivered.

One of the big things now is with ever-increasing imaging you've got an ever-increasing burden of reviewing...If AI allows you to do that at a more efficient pace and a more accurate pace. [P2]

Six optometrists expressed concerns that colleagues using these tools might fail to develop critical skills.

If it’s brought in too early in somebody’s training as a clinician, then [they may] fail to develop their own clinical decision-making skills. [P18]

There was a similar concern that skills might fade among those who become too dependent on the new technology, coupled with a risk of misinterpreting “results in the AI.”

We’d have to be very clear that whoever’s [using] the AI understands how to interpret them, as well as preventing skill fade from clinicians who are used to interpreting these images. [P2]

All optometrists but 1 were enthusiastic about taking on new responsibilities to improve patient care. This would require additional skills and professional development. P6 observed that it might create a rift in their community.

There are some practitioners...a small minority...they were trained to examine eyes and provide optical corrections and they don’t like this whole shift. [P6]

Five optometrists feared that reliance on AI assistance would impinge upon their skills and professional judgement and their personal contact with patients.

There is a risk of reducing the respect, qualifications, and the ability of the optometrist...[AI] is used to replace parts of a test rather than aid. [P10]

However, many optometrists considered that new technologies will improve their reputation (7/18) and help them be seen as up-to-date (1/18).

If you’re explaining to a patient that you’re using AI, they would be very impressed...happy...their optometrist is using up-to-date methods. [P4]

There is a considerable delay from the moment AI research demonstrates a new technique to its wide application. One optometrist expected that the national repository will “reduce this time of ‘translating’ research into practice to 10 years” (compared with 17 years [6]) and “hopefully save some eyesight for people” [P3].

**Theme 4: Paths to Improved Eye Care**

Taking a long view of improvements in eye care made possible by the image repository and AI, optometrists expected early detection (18/18), increasing accuracy (10/18), higher efficiency (4/18), better disease progress monitoring (8/18), and risk prediction (7/18). Two research advances using the repository are anticipated: (1) improved AI decision support (tuned for the population and with new predictors) and (2) improved education. These are assumed by optometrists when they discuss long-term benefits.

We may be able to catch things before they get to a more progressive stage where they are more devastating to sight as well. [P13]

Earlier accurate diagnoses would increase optometrists’ efficiency.

Often, we’ll see patients...for follow-up appointments just because we’re uncertain. But if these AI technologies...even in those grey or uncertain patient situations meant we could make a better or a quicker judgement, then that would be handy. [P18]

The speed of diagnosis increases efficiency and reduces patient stress.

It needs to be as close to real time as possible so that it can be a very clear way of generating a result that doesn’t have the anxiety of waiting on an envelope coming through the door or an email. [P2]

Support for managing disease progression is a widely held expectation.

We need to know whether it’s the same pathology, whether it’s...progressed. [P27, ophthalmologist]

Optometrists expect to be in a better position to predict risks.

Then we would be in a better position to predict patients at risk. Rather than monitor patients who already are showing signs of disease...It may be easier to predict patients at risk of developing certain conditions. [P10]

Sustainability was a concern for 6 optometrists. Some suggested that the government should support not just the acquisition of new equipment but also the organizational procedures and systems they needed.

In Scotland, the funding from Government would have to match the amount of time that’s actually spent using this [AI] equipment to aid with a diagnosis. [P6]
Theme 5: Education and Training

All interviewees agreed that education is needed to prepare them to exploit image sharing and AI tools in a reliable and professional way. The main points revealed were (1) formalizing and incentivizing training, (2) information resources, (3) allocating the time for training, and (4) learning by interacting with ophthalmologists.

Five optometrists suggested developing accredited educational resources about AI and its role in eye care to encourage participation and to validate achievement of standards.

...if all optometrists had to do mandatory training as part of their CPD [Continuing Professional Development]. Part of our mandatory training, then we would all be starting on the same page. [P15]

The education of patients and their supporters was perceived by optometrists as crucial. It takes time to discuss new approaches. They needed information that was simple and concise in various formats (eg, on paper, online).

...You’d certainly need something...in terms of cascading the information to patients...an information pack [we] can just hand out. [P6]

Steering groups were proposed to raise patients’ awareness of the benefits of AI, discuss concerns, and clarify challenges.

...You could have a steering group with optometrists, ophthalmologists, tech guys, patients represented, to hear their views, but maybe you’re doing that already. [P6]

There was an almost universal feeling that the best way to develop the new skills needed was by learning on the job. Several suggestions emerged as interviewees contemplated what might be needed when a small number of ophthalmologists were helping a larger number of optometrists develop their professional expertise. Jointly developed treatment plans were an aspiration for 5 optometrists and 2 ophthalmologists to improve their relationships and train optometrists for more roles.

Two ophthalmologists suggested having more optometrists visiting hospitals.

...I’m very [keen to have] hospital optometrists working in our teams. I feel that I can support the training, development and progress of these small cohorts very well rather than communicating with multiple community opticians. [P22, ophthalmologist]

Discussion

Optometrists deliver primary eye care service for the great majority of patients and judge when referral to secondary care is warranted. Our study has established a model for enquiring about their aspirations and concerns. We summarize the wide-ranging discussions, which thematic analysis clustered into 5 themes that interlink.

Theme 1: Changes to Professional Working Patterns

Changes to the professional working patterns of optometrists are anticipated due to increasing clinical responsibilities and growing workloads. These stresses arise from an aging population with increasingly severe eye conditions and effective treatments extending the duration of care. Extra information provided by an increased number of higher resolution images requires additional interpretation time. Funding was optometrists’ primary concern, with refractive correction a potential source of cross subsidy in prosperous practices. A variety of ways of providing more care in the community anticipated their discussion of theme 4.

Theme 2: Envisaging the Image Repository’s Impact

Optometrists looked toward (1) professional collaboration, (2) teleconsultations, (3) remedies for health inequalities, (4) image quality issues, (5) effective standards and interworking systems, (6) scrutiny of decisions, and (7) EHR for optometry. The benefits of covering the full diversity of patients and eye conditions predominated. Optometrists’ vision went beyond the direct effects of an image repository. For example, after the pandemic’s restrictions, they envisaged triage improvements from image-sharing teleconsultation. They expected to support colleagues working in deprived communities, and they foresaw an integrated image-handling EHR system improving their management of patients. OpenEyes, being commissioned in Scotland, will meet EHR requirements, but community optometrists will have to wait for its benefits, as it will be deployed initially in hospitals [14]. These extensions and worries over scrutiny reveal misconceptions about the SCONe repository. Its privacy protection extends across all patients and optometrists; so, neither data sharing nor scrutiny are possible within SCONe. When an EHR provides those mechanisms, these opportunities and issues will re-emerge.

Theme 3: Benefits From AI Decision Support

In this context, optometrists expect (1) more accurate and faster decisions for which they would still take full responsibility, (2) help with conditions not previously encountered, (3) more efficiently interpreting images, (4) short-term status enhancement with patients despite some fears of longer term erosion of expertise and responsibility, and (5) frustration over delays in adopting new methods leading to loss of sight that could have been prevented. Invariably, AI assistance was anticipated positively but with significant concerns about responsibilities, practicalities, and time scales. The support for the image repository and the AI it enables could evaporate unless the pragmatic issues raised regarding system and technical complexities, the impact on already busy workloads, and the navigation of ethical and patient privacy governance are addressed. González-Gonzalo et al [6] recommend a multistep strategy engaging all stakeholders to address this. This careful planning and introduction process is necessary to prepare for provision and to sustain such innovations. There remain uncertainties about the effectiveness in the field of AI-powered decision aids in the context of evolving practices, diversity of equipment, and variations in image-taking procedures inherent in community optometry.

Theme 4: Paths to Improved Eye Care

Optometrist suggestions included (1) earlier detection of more conditions, (2) increasing triage accuracy, (3) improved efficiency, (4) better condition monitoring, and (5) identifying
patients at risk. The earlier detection depends on research enabled by their contribution of earlier images (and similar research) detecting latent signals. Improved triage accuracy depends on the AI assistance and the additional training both enabled by the repository. Studies of whole populations enabling AI assistants to highlight signals they might otherwise miss would underpin these benefits. Interviewees anticipated closer collaboration between primary and secondary care, redistributing responsibilities for diagnosis and treatment for some disorders, leading to new roles for optometrists with additional skill requirements. However, there was concern among optometrists that communication from hospital eye services to optometrists needs improving. An audit highlighted variability (45%-92%) in the successful delivery of formal responses from hospitals to optometrists [3]. Sustainability is a critical issue, involving many more professional roles [6].

**Theme 5: Education and Training**

In this context, optometrists proposed (1) incentives via certification, (2) informing patients and supporters, (3) ophthalmology placements, and (4) jointly planned treatments. All interviewees valued training and expected significantly informative extra material drawn from the repository. However, contributing to the repository, using the AI assistants, and taking on extra clinical responsibilities will all require additional training, requiring more resources, time, and materials.

A cross-cutting issue emerged. The difference in perception of the quality of communication between optometrists in the primary care sector and specialist ophthalmologists is worrying and merits further attention. It may result from differences in professional status (eg, reflecting different lengths of qualification path [4 and 7+ years]) as well as the different relationships between professionals and patients (patients are free to switch between optometrists). Improvements in digital communication (EHR) may mitigate or exacerbate this issue.

In the evolving primary eye care context revealed by theme 1, themes 2 and 3 meet our objective of discovering optometrists’ attitude to contributing their patients’ retinal images and to using AI assistance. Theme 4 captures their ideas about how eye care may be improved by future service innovations—our second objective. These require training innovations covered by theme 5. As few new ideas emerged in the final interviews, we regard our methodology and evidence gathered as reliable.

Primary eye care in Scotland depends on the skills and diagnostic capabilities of optometrists developed and assessed through the NHS Education and Glasgow Caledonian University [7]. The local culture provides a positive context for these developments. However, we believe that in most other contexts, optometrists would have similar aspirations and concerns. Our methodology and findings should prove beneficial for other countries planning to gather optometry images to improve early detection and triage. Further work should explore these developments in an international context considering all the issues optometrists raised.

**Limitations**

As a pilot study, we only interviewed 18 out of 1300 optometrists. Our very small sample has a potential bias that is hard to estimate, as interviewees were recruited through the SCONe project. Therefore, respondents may have been better informed and more inclined to support retinal image contribution. Table 2 shows that we achieved reasonable coverage of geographic diversity.

**Acknowledgments**

We greatly appreciate the time and effort contributed by all our interviewees. We greatly appreciate the criticisms and suggestions of JMIR Publications reviewers and 3 internal reviewers, namely, Jano van Hemert, Rob Proctor, and Philip Welsby. This work was supported by the Scottish Funding Council as part of the Edinburgh City Initiative administered by the Data-Driven Innovation center (grant GZ9261). We thank the whole Scottish Collaborative Optometry-Ophthalmology Network e-research team supported by the Chief Scientist’s Office, Edinburgh and Lothians Health Foundation, Sight Scotland, Royal College of Surgeons of Edinburgh, and RS Macdonald Charitable Trust.

**Conflicts of Interest**

None declared.

**References**

2. Scottish Collaborative Optometry-Ophthalmology Network e-research. URL: https://www.ed.ac.uk/clinical-sciences/opthalmology/scone [accessed 2022-12-21]


11. QSR International Pty Ltd. URL: https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home [accessed 2022-12-21]


14. OpenEyes. URL: https://openeyes.apperta.org/ [accessed 2022-12-21]

**Abbreviations**

AI: artificial intelligence
EHR: electronic health record
NHS: National Health Service
SCONe: Scottish Collaborative Optometry-Ophthalmology Network e-research

©Aurora Constantin, Malcolm Atkinson, Miguel Oscar Bernabeu, Fiona Buckmaster, Baljean Dhillon, Alice McTrusty, Niall Strang, Robin Williams. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 25.05.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Patients With Cardiovascular Disease Revisiting Specialist Physicians via Remote Treatment: Interview Study of Experiences

Charlott Ek¹*, RN, MSc; Per-Daniel Liljegren²*, EdS, MSc; Anette Edin-Liljegren³*, BSc, PhD

¹Jokkmokks Healthcare Centre, Region Norrbotten, Jokkmokk, Sweden
²Primary Care, Region Västerbotten, Umeå, Sweden
³Department of Epidemiology and Global Health, Umeå University, Umeå, Sweden

*all authors contributed equally

Corresponding Author:
Anette Edin-Liljegren, BSc, PhD
Department of Epidemiology and Global Health
Umeå University
Byggnad 5B, 3-4th Fl Norrlands universitetssjukhus
Försörjningsvägen 7B
Umeå, 901 87
Sweden
Phone: 46 706457197
Email: anette.edin-liljegren@umu.se

Abstract

Background: Access to health care for an aging population with growing needs presents major challenges in northern Sweden’s sparsely populated regions. Few people, the lack of professionals, and long distances make it difficult to provide health care on equitable terms according to the Swedish legislation. Remote treatment (RT) using information and communication technology has been suggested to overcome these difficulties, and person-centered care (PCC) is a desired philosophy to improve the quality of health care. However, there is scarce knowledge about how patients experience RT meetings.

Objective: This study aimed to describe the experiences of patients with cardiovascular disease revisiting specialist physicians via RT guided by a PCC perspective in northern Sweden’s sparsely populated regions.

Methods: A qualitative study was conducted based on interviews with 8 patients with cardiovascular disease revisiting their physician through RT, from a digital health room to a health care center or from a health care center to a hospital. The interviews were recorded, transcribed verbatim, and analyzed using inductive content analysis. The results are discussed from a PCC perspective.

Results: The analysis resulted in 6 categories: good accessibility, safety with good relationships, proximity and distance with technology, habit and quality of the technology facilitating the meeting, cherishing personal integrity, and participation in own care. These categories were interpreted as the theme, participation and relationships are important for good and close care via RT.

Conclusions: The study shows that participation and relationships are important for good and close care via RT. To improve the quality of an RT meeting, PCC can be applied but needs to be extended to the digital domain—electronic PCC, especially the communication component, as it is the most salient difference from a face-to-face meeting. Important factors that should be considered before, during, and after the RT meeting have been identified.

(JMIR Hum Factors 2023;10:e43125) doi:10.2196/43125

KEYWORDS
remote treatment; sparsely populated region; telemedicine; content analysis; experiences; person-centered care; rural; eHealth; mobile phone
Introduction

The Challenges of Health Care in Sparsely Populated Regions

Sparsely populated regions (SPRs) have many things in common across the world such as low population density, long distance to health care and other societal services, being governmentally remotely controlled, and a distinct lifestyle [1]. These regions are not only different from urban areas but also different from rural regions in general and have been described as a specific geographic category comprising >60% of the Earth [1]. Therefore, it is important to study access to health care under these conditions. In the SPR of northern Sweden, demographic transition and urbanization have led to a large proportion of older adults still living in their homes [2]. The geographical location with long distances to health care units makes access to care challenging [3-5], and studies show that people living in SPRs receive poorer care than those living in cities [5], which is contradictory to Swedish law where “the goal of healthcare is good health and care on equitable terms for the entire population” and the care should be organized close to the people [6]. There have been several highly prioritized initiatives from the Swedish government [7] to ensure good-quality, local health care; however, it is still unclear how this should be implemented in SPRs.

Opportunities With Digital Technologies

A way to overcome these challenges is to use information and communication technology (ICT), which has been recommended and encouraged by the World Health Organization [8]. An initial statement from the World Health Organization Bellagio eHealth Evaluation Group proposed that “To improve health and reduce health inequity, rigorous evaluation of eHealth is necessary to generate evidence and promote the appropriate integration and use of technologies” [8].

It is important to evaluate the implemented methods and techniques because despite its many benefits, the introduction of new technology may lead to new problems, such as patient integrity and safety issues [9]. In a Danish Island, more than half of the patients did not like consulting a specialist via ICT [10]. In the study, a large proportion of older adults and people with only primary education indicated that there could be difficulties in introducing ICT in rural areas where the level of education is lower, in general, than in urban areas. In a study by Call et al [11], overall, 43% of the participants were still averse to telemedicine despite the inconvenience of in-person visits. Recently, the COVID-19 pandemic has further enforced the use of telemedicine [12]. However, there is both a lack of consensus of terminology and a knowledge gap regarding how various aspects of telemedicine work.

The level of education, previous use of social media and other communication platforms, and being a rural resident are factors that may influence how receptive the informants are to telemedicine [10,11]. Age is correlated negatively to computer literacy [12,13], which affects the outcome of the introduction of technology. In contrast, several studies have shed light on the importance of telemedicine from the perspective of patient satisfaction [5,14-16]. Patients reported saving time and reducing costs by not having to travel and were satisfied with the technical performance [17-20]. Furthermore, some patients who were negative about using video meetings initially changed their minds when they tried it [15]. Therefore, there is reason to believe that follow-ups of planned care visits via ICT could be a valuable complement to physical meetings when physicians and patients have already established a relationship.

To provide more qualified care for people living in SPRs, remote treatment (RT), which we define as treatment that is conducted remotely by means of ICT, including medical advice, examination or treatment, where the patient and therapist are separated in space, but not in time, may be an option. Thus, we considered RT as a subset of the broad concept of telemedicine to limit and clarify the aim of this study [21].

RT is also important for sustainable health care and is likely to be of great benefit for patients, professionals, caregivers, and society and is a way of increasing accessibility to health care on equitable terms and supplying specialized skills to remote areas. Therefore, there is a great need for systematic studies to ensure the quality of RT meetings and to obtain patient experiences of safety, partnership, and shared decision-making. Thus, RT can save both time and money, primarily for patients who must travel long distances to health care units, and can also reduce the carbon dioxide footprint of health care [3,22].

As cardiovascular diseases are among the most common diseases in the world and in Sweden [23], it is important to increase knowledge about these patients’ experiences of seeing their physician for planned follow-up meetings via RT and how telemedicine solutions can be a way to increase access to health care in SPRs.

Person-Centered Care—A Desired Model

Person-centered care (PCC) is a care philosophy that aims to include the life-world perspective and seeing the whole person and has been developed to improve the quality of health care [24,25]. A transition from a care model with the patient being seen as passive to being active in their own care and their own resources used are important factors [24]. PCC creates a sense of self-empowerment to manage one’s own illness; contributes to safety; and is linked to short care times, fewer readmissions, and better quality of life for the patients [26]. PCC is a collaboration and a partnership between the health care staff and the patient. It is a mutual approach in which health care professionals respect the knowledge that the patient can provide about their own life and health situation, such as values, goals, and previous experiences. The health care staff contribute with their professional expertise and information about care alternatives [26].

PCC means that “individuals’ values and preferences are elicited and expressed, guide all aspects of their health care, supporting their realistic health and life goals” and is achieved through a dynamic relationship between individuals and health care professionals [27]. Recently, PCC has been proposed as a desired care model in several Swedish governmental reports, which is a step toward legislation [7,23,28].
The rapid development of digital technologies and the need for transformation of the health care system make PCC a natural starting point for investigating RT.

Thus, both RT and PCC have been suggested to improve the accessibility, efficiency, and effectiveness of health care [7,24,29]. However, there is still a lack of knowledge about how patients experience RT meetings.

This study aimed to describe the experiences of patients with cardiovascular diseases regarding follow-up meetings with their physician through RT, in northern Sweden’s SPR, guided by a PCC perspective.

**Methods**

**Study Design**

A qualitative approach was used to reflect the experiences of people who receive RT. The data were originally collected in a master thesis at the Department of Nursing at Umeå University and were further analyzed in this study. According to the guidelines for necessitating quality and transparency of health research, Consolidating Criteria for Reporting Qualitative Research (COREQ) [30] were followed during the process.

**Participants and Settings**

This study was conducted in the SPR of northern Sweden, where Region Västerbotten and Region Norrbotten are official health care providers. RT was conducted with a patient and a specialist physician having a digital meeting between a health care center (HCC) and a hospital—or between an HCC and a digital health room (DHR; Figure 1).

The DHR is a room equipped with ICT, an encrypted videoconferencing system, and other medical devices that are not available at home. The DHR has been established in small villages in Västerbotten County, close to the inhabitants, to provide more accessible and equitable health care. In this remote area, the distance to the nearest hospital could be >300 km. Before and during some of the meetings, the staff was sampling, performing examinations, and supporting the patients with connection to the videoconference system. A digital stethoscope was used to transmit heart and lung sounds in real time to the connected medical specialist during the visit for some patients. The operation manager from the HCC and a nurse at the hospital recruited participants for this study. The inclusion criteria were patients aged >18 years diagnosed with cardiovascular disease who have had a planned revisit to their physician via RT. Participants were informed in writing and orally about the study and asked whether they would participate, and they signed an informed consent form before the interview started. The 8 participants consisted of 4 (50%) women and 4 (50%) men, aged 53 to 85 years (Table 1).

**Figure 1.** Settings for remote treatment.
Table 1. Patient characteristics and way of connection to specialist physician from a digital health room (DHR) to a health care center (HCC) and from an HCC to a hospital.

<table>
<thead>
<tr>
<th>Patient ID</th>
<th>Age (years)</th>
<th>Sex</th>
<th>RTa connection</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>83</td>
<td>Female</td>
<td>DHR to HCC</td>
</tr>
<tr>
<td>2</td>
<td>85</td>
<td>Male</td>
<td>DHR to HCC</td>
</tr>
<tr>
<td>3</td>
<td>81</td>
<td>Male</td>
<td>DHR to HCC</td>
</tr>
<tr>
<td>4</td>
<td>72</td>
<td>Female</td>
<td>HCC to hospital</td>
</tr>
<tr>
<td>5</td>
<td>75</td>
<td>Male</td>
<td>HCC to hospital</td>
</tr>
<tr>
<td>6</td>
<td>68</td>
<td>Female</td>
<td>HCC to hospital</td>
</tr>
<tr>
<td>7</td>
<td>78</td>
<td>Male</td>
<td>HCC to hospital</td>
</tr>
<tr>
<td>8</td>
<td>53</td>
<td>Female</td>
<td>HCC to hospital</td>
</tr>
</tbody>
</table>

aRT: remote treatment.

Data Collection

The data collection was inspired by the PCC philosophy, but because there is no unified theory [25], we constructed open-ended questions [31] about people’s experiences of meeting their physician at a distance. Data were collected using semistructured interviews by the author (CE). The questions were about what worked well and what did not in the digital meetings, relations, experiences of connection to the physician, differences between physical and digital meetings, what it means to get access to digital meetings, and how the meeting could be improved. Each question was followed by further questions to develop previous statements and encourage the interviewee to talk more about the situation and give examples. The interviews were conducted at the participants’ homes, recorded digitally, and transcribed verbatim. As it was difficult to find participants, the study and the interviews were conducted over 2 periods—during July 2017 and from November 2018 to February 2019; the interviews lasted 22 to 42 (median 29) minutes.

Data Analysis

The transcribed text was analyzed systematically using content analysis described by Graneheim and Lundman [32]. The results were then interpreted inductively, which is recommended in the literature if the knowledge gap of what will be studied is limited or fragmented [33]. Furthermore, the interviews were read through carefully several times to get a sense of the whole of the material. Text units with the corresponding purpose of the study were chosen and condensed. The text units were coded close to the text, which were then abstracted into subcategories. Furthermore, subcategories that were similar to each other were sorted and abstracted into categories. The categories related to each other, and the underlying sentences were interpreted and formulated in a theme such as descriptions as a common thread, where the sentence reappeared in category after category [33]. Overall, 2 authors (CE and AE-L) discussed codes, subcategories, categories, and the theme with each other during the analysis process to ensure the credibility of the study [34,35].

Ethics Approval, Informed Consent, and Participation

The study was conducted in accordance with the ethical principles described in the Declaration of Helsinki [36]. In SPRs, one needs to be careful about ethical issues when presenting data about patients. In a small village, even age and sex in combination with a medical condition may be sensitive data for identifying a person. Informed consent was obtained from both the operation managers and the participants. Participants were informed both in writing and orally about the possibility to participate in the study and that they could cancel their participation at any time without providing any reason [36]. Participants were also informed that personal information and data from the interviews could not be attributed to the individuals and that they have been treated confidentially. The study was approved by the Regional ethical review board located in Umeå (2017/155-31 and 2018/237-32).

Results

Overview

The analysis of the interviews resulted in 6 categories: good accessibility, safety with good relationships, proximity and distance with technology, quality of and familiarity with technology facilitating the meeting, and cherishing personal integrity and participation in care. The categories were abstracted and sorted from a total of 16 subcategories, as shown in Textbox 1. From the categories, a theme was interpreted as participation and relationships are important for good and close care via RT. The categories and the theme are presented in the following sections and illustrated with quotes from the interviews.
**Textbox 1.** Overview of categories and subcategories of the theme participation and relationships are important for good and close care via remote treatment.

<table>
<thead>
<tr>
<th>Good accessibility</th>
<th>Traveling and safety</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time saving</td>
</tr>
<tr>
<td></td>
<td>Equitable care</td>
</tr>
<tr>
<td>Safety with good relationships</td>
<td>Familiarity with the staff</td>
</tr>
<tr>
<td></td>
<td>Comfortable togetherness in the waiting room</td>
</tr>
<tr>
<td></td>
<td>Relatives provide support</td>
</tr>
<tr>
<td>Proximity and distance with technology</td>
<td>Being close and feeling distance via the video screen</td>
</tr>
<tr>
<td></td>
<td>Personal and impersonal contact</td>
</tr>
<tr>
<td></td>
<td>Calm and focused meeting</td>
</tr>
<tr>
<td>Quality of and familiarity with technology facilitating the meeting</td>
<td>Being familiar with the technology makes the meeting easy</td>
</tr>
<tr>
<td></td>
<td>Supported or disturbed by technology</td>
</tr>
<tr>
<td>Cherishing personal integrity</td>
<td>No public self-disclosure</td>
</tr>
<tr>
<td></td>
<td>Importance of privacy</td>
</tr>
<tr>
<td>Participation in care</td>
<td>Being prepared</td>
</tr>
<tr>
<td></td>
<td>Wanting more information</td>
</tr>
<tr>
<td></td>
<td>Opportunities for development</td>
</tr>
</tbody>
</table>

**Good Accessibility**

The informants expressed that it was valuable to reduce the time spent in traveling to revisit their physician and to increase safety by not having to drive. The possibility to meet the specialist via RT was perceived as more equitable care and was interpreted as a common category, *good accessibility*.

**Traveling and Safety**

Participants in the study experienced a great advantage in avoiding traveling, thus reducing the amount of driving. It was convenient and easy with the short route, or the health care unit was so close that the informants could walk to the meeting instead of traveling long distances and seeing the physician for just a short time. Although the interviewees living in SPRs were used to traveling long distances, in winter, with difficult and unpaved roads, it was especially valuable to avoid traveling. The individuals also experienced that the evening sun in the eyes could be tiring when driving. Owing to the northern location being close to the Arctic circle, during the winter season, the sun is very low to the east in the mornings and very low to the west in the afternoons and thus in the eyes—both ways to the HCC unit and home:

...I thought it was great because then you do not have to go to [the hospital] and get away from driving and all that...and because it is so close it is only a couple of minutes to walk there... [ID6; female; aged 68 years]

Some informants found it difficult to drive far owing to illness and pain. Reducing travel also meant an economic advantage, as they used their own car to drive to the HCC or hospital:

...Say that it cost 150–200 SEK in fuel for a trip and then you get 24 SEK for it in compensation...it is not worth it... [ID8; female; aged 53 years]

Even when it was easy to go to the health care unit, the taxi ride to the hospital or HCC could be agreeable with drivers you know. Despite long distances, the journey could be pleasant, and sometimes, it was not difficult to travel, especially when informants also took the opportunity to go shopping or do other errands at the same time:

Even so, a trip to [the city] means you can go to the shops before and after and do errands there. [ID6; female; aged 68 years]
Time Saving
Some participants indicated that having the meeting with the specialist via RT saved time and that they received help more quickly and avoided worries. RT made the whole day easy and took just 15 minutes compared with the fact that it takes half a day to see a physician in a hospital. Participants felt that time was saved for both themselves and the physicians. Time was also saved for the staff, whose job is to support with connections and keep track of the routines so that the physician’s visit was not delayed. An interviewee said that everyone has the same amount of time and we live in a stressed society; therefore, it was good that care could be provided remotely for people with long distances to health care units. A participant of working age saw benefits in not having to take time off, not missing working hours, and earning income:

And if you work and...I do not have to miss so much working time and do not keep on and may not need to compensate for work so much and make changes with colleagues...so there is not much lost work income either...there are financial benefits... [ID8; female; aged 53 years]

Equitable Care
Participants in the study felt that the care was equitable and that they received the same assessment as at the HCC or at the hospital for this type of revisit. The RT meeting felt normal and was not different, except that the physician and the patient were not physically in the same room. If they had gone to the HCC or to the hospital to meet the physician face-to-face, the physician would have asked the same questions as asked during the digital meeting:

The great thing is that it is equitable...what should I say...it gives just as good results with these technical facilities [video]. [ID2; male; aged 85 years]

Safety With Good Relationships
In the interviews, it was noted that the patients were familiar with the staff and experienced a comfortable time together in the waiting room, including relatives who gave support in the meeting with the physician, which was interpreted as safety with good relationships.

Familiarity With the Staff
Participants in the study were familiar with the staff at the health care unit and already knew the physician before the meeting, which created security and a feeling of safety in the meeting. The informants thought that the physician was pleasant and easy to talk to during the meeting. A participant knew his physician only through phone before but felt that the meeting worked very well. It was very important that they had met the physician before, instead of meeting a new unknown physician who did not know anything about them as a person:

But had it been a complete stranger then you get a little...then you keep a little distance...if you...you think what the heck is this. [ID7; male; aged 78 years]

Comfortable Togetherness in the Waiting Room
A good social gathering was experienced when the patients met acquaintances in the waiting room at the HCC unit. It became a pleasant meeting place, similar to going to the neighbor’s house and meeting people you know. The conversations were relaxing, and there were discussions about what had happened since the last time they had met, how it was on the fishing trip, and even some talk about illness. A participant thought that it felt similar to home and he could be himself:

...We were standing out there talking and then we entered and then you could have coffee if you wanted and another acquaintance was sitting there and it was no big deal to get there as dressed as when you walk in a village, it feels like home in some particular way, yes... [ID1; female; aged 83 years]

Relatives Provide Support
Close relatives could be a great support in the RT meeting, which provided a feeling of safety both before and during the meeting. Participants experienced that close relatives were a support when the patient had hearing or memory problems, for example, after a stroke:

And so I had [the man] was there to support me if I forgot something or if I lost words... [ID6; female; 68 years]

Proximity and Distance With Technology
Both proximity and distance were experienced in the RT meeting, a contact that could be perceived both as personal and impersonal. The meeting felt calm and focused on the patient themself, which was interpreted as proximity and distance with help of technology.

Being Close and Feeling Distance via the Video Screen
The image on the screen was large and clear, and the patients were affirmed by the physician on the screen. Participants felt that it was as close as in real life, almost similar to sitting in the same room. Some informants perceived it as if the physician was behind the video screen and that they had eye contact. The patients experienced that the physician could see their reactions and facial expressions:

That he or she can look at my face and could see how I think before I answer, I actually think if I’m honest or making up [laughter]...that’s exactly what I think I can do with the grandchildren when I talk to them on Skype. [ID1; female; aged 83 years]

The interviewees experienced that the sound was good when speaking to the physician, without interruption, and that it was easy to communicate. A person described how they went through the medication list in the RT meeting. The patients were also impressed by the possibility that the physician could zoom in on them and listen to their heart and lungs through the digital stethoscope:

They listen to my heart and then hear all that, 75 km away, so the physician can sit and listen to my heart, it’s so amazing it’s crazy. Ahh...I think that’s really impressive. [ID1; female; aged 83 years]
Also, a feeling of distance was experienced by some informants in the RT meeting. It felt different, and some participants found it difficult to be spontaneous and answer the physicians’ questions. For them, it was difficult to see body language and facial expressions:

*If I talk to a person sitting in front of me, I can see their body language, I can joke with the person...I can ask and say things that are almost private but a person who is on a screen is a bit distant because I feel like I can’t really talk.* [ID 6; woman; aged 68 years]

The feeling of absence of physical contact was experienced when the physician could not touch the patient and measure the pulse. They thought that physical contact should be the right way to meet the physician, because people become more sensitive if they are close to each other. Some patients were still satisfied and thought that the meeting was normal without physical contact and that the on-site nurses could do the examination. However, a participant wanted a physical meeting:

*They never asked me what I wanted they just said it would be through video but I would have preferred to meet them there [at the hospital].* [ID4; female; aged 72 years]

**Personal and Impersonal Contact**

Patients in the study found the personal contact to be perfect even though they met the physician via a computer screen and they were not physically in the room:

*It was like personal contact even though it was via such a link.* [ID7; male; aged 78 years]

Some patients felt the opposite, that the physician’s visit was impersonal via RT, and they experienced a feeling of insecurity. Some other participants had problems in getting something out of the meeting; they did not ask their questions because the meeting did not feel personal. A person wanted to meet the physician physically because he had vision and hearing problems, which resulted in the physician feeling like a stranger in the RT meeting. It became uncomfortable; therefore, the person barely remembered the meeting and thought it was something wrong with her but said that the physician was certainly professional:

*I would rather have a personal meeting, you can reach them in other ways when you have vision problems and sitting close because I like to comment on things under...and when they were like strangers to me, I couldn’t.* [ID4; female; aged 72 years]

Another participant expressed that although the meeting did not feel personal, it worked to meet the physician via the video screen:

*It’s not this kind of personal, so I feel, but I thought it worked well.* [ID7; male; aged 78 years]

**Calm and Focused Meeting**

The RT meeting was experienced as calm and focused and almost as in a home environment. The patient got easy contact with the physician in a peaceful and quiet way, and nothing was disturbing from the background. The informants thought that the physician was responsive and gave them time to ask questions. This meant that the patients did not feel stressed and felt that the physician was focused on them during the meeting. At an ordinary physical meeting with the physician, people look around at things on bookshelves and other things in the room, but all that disappeared in the RT meeting, which was perceived as positive. A person got the feeling that the meeting was focused as only one could talk at a time:

*I don’t know if it was because of the technology...You have to be quiet, it felt that way anyway...When one talks, you listen to what he will say...But I don’t think it’s something negative...then you get even more focused than maybe talking at the same time...because then neither of us really listens.* [ID8; female; aged 53 years]

**Quality of and Familiarity With Technology Facilitating the Meeting**

The informants experienced that technical skills facilitating the RT meetings but that technical quality could both support or disturb the meeting, resulting in the interpretation of the category, quality of and familiarity with technology facilitating the meeting.

**Familiarity With the Technology Makes the Meeting Easy**

The interviewees felt that the physicians gave the impression that they were comfortable with the technology. In addition, the patients themselves felt comfortable. It was not strange because they were used to the technology related to using the internet, Skype, or other systems through their work:

*But you’re used to watching TV so you’re not completely alienated from things like being on Skype with grandchildren on the iPad.* [ID1; female; aged 83 years]

Some participants also felt unfamiliar and insecure when they had a digital meeting. They knew it was possible to meet the physician via RT but had not had any meetings themselves before. The first time felt special, strange, and stiff because everything was new but, at the same time, exciting. The informants experienced the feeling of not having control, but after a while, they got used to it:

*And then the physician came, and it felt a little bit strange when you’re not like this [Physically]...but you see a TV screen, but after we had been sitting for a while there was nothing strange about that.* [ID7; male; aged 78 years]

There were participants who felt old-fashioned when they would talk via a video screen, but they thought it was a matter of age and habit. Participants hoped for more opportunities to have RT meetings, and they imagined that when they got used to it, it would feel similar to sitting in front of a physically present physician. In addition, a person felt that the physician was uncomfortable with the technology:
Supported or Disturbed By Technology
Some interviewees told us in the interview that they got support and help from the staff to start the meeting. The informants also mentioned that when they arrived at the HCC unit, they were directed to a room with a table and a video screen, and the staff started the computer and instructed them about how to use it. In the RT meeting some participants also mentioned how the physician instructed the staff about how to put the stethoscope in place to be able to listen to the sounds from the heart and lungs. An informant said that the physician informed them how the videoconference would be conducted:

...She told me where it was and that she was going to ask me a few questions...And I said it's just to ask questions...I'll answer as best I can. What I understood, it went as well as possible. [ID5; male; aged 75 years]

Overall, the technology worked well, but participants felt that the technique could be disruptive. A participant said that it was a hassle with the sound and it was difficult to hear the lung and heart sounds using the digital stethoscope, but it started to work at the end of the meeting. Another participant could only see half of the physician’s face and thought it felt strange:

...At last she got herself on the screen but it just happened that we only saw half of her, I think we saw her from the nose and upward so she sat like in a corner of the picture, and then we talked to her but it felt quite strange to sit and talk to half a person, half a face. [ID6; female; aged 68 years]

Cherishing Personal Integrity
The patients felt that they did not want to disclose information about themselves at the digital meetings and that it was important to have a private room when they met the physician, which was interpreted as cherishing personal integrity.

No Public Self-disclosure
Participants in the study said that several individuals were present during the RT meeting and that they did not want to disclose themselves to people other than the physician. Other people could be present because the health room was also used as a gathering point for the home care service in the area. Participants thought that it was a sensitive situation, and they did not want other people to hear what thoughts, worries, and illnesses they had, even though they knew the staff had a duty of maintaining confidentiality:

I may not want so many people listening and hearing what I’ve been thinking about, what illness, what thoughts and what problem I have or concern... [ID1; female; aged 83 years]

Importance of Privacy
Participants pointed out the importance of individual meeting rooms and that people should not pass by all the time. When people passed by, it was difficult to focus on the meeting and maintain confidentiality:

Then you sat alone in your own secluded room and it is also quite important. [ID8; female; aged 53 years]

Participation in Care
The informants had a desire to be prepared for the meeting with the physician, wanted more information and influence, and saw opportunities for the development of care. These are summarized in the category participation in care.

Being Prepared
Participants had written down questions and thoughts they had—such as medications, how they would think ahead, and future follow-ups—and they wanted to be prepared for the RT meeting. It was important to be prepared; otherwise, they had the risk of forgetting half of their thoughts:

I probably got answers to all the questions I had; I had written down what to ask for and she answered... [ID6; female; aged 68 years]

Wanting More Information
There was a request from the participants for more information before the RT meeting. As first-time users, they had heard of an appliance they could talk to and thought that someone would be there to give support and tell them how it worked practically. A participant was concerned that the screen was not switched off after the visit:

I felt awfully bad because I thought now it’s on...what if...A lot of these, what if...and standing in [the hospital]...and what if the power is on? All that practical stuff...I walked around and thought about it for a long time. [ID6; female; aged 68 years]

There was a desire for information before the appointed meeting, so that everything could be arranged properly. The participants wondered whether there were any routines at the HCC unit when they got the feeling that no one knew anything about the meeting, for example, who shows the patient the way and who would initiate the meeting via the video screen. A person lacked information and felt that she was not involved in the visit, and she wanted someone to coordinate the visit:

But now I understand that it’s an expense for the healthcare system but in this particular case it would only have been the cost of one trip that I would have anyway when I was going to [the hospital]. [ID4; female; aged 72 years]

Opportunities for Development
Joy and hope for the future were expressed by the participants. Some informants experienced the meeting as fantastic and wished to continue conducting RT meetings—not only for people who lived in SPRs; however, it was people in SPRs who made the most of such visits. It was important that the physician should be known to the participants before the RT meeting and could communicate understandably. Interviewees in this study could see that there were opportunities to develop digital meetings, but it was clear that the need to go to the hospital in
more difficult cases was obvious. Developing remote care for mild ailments was welcome. If the physician wanted to do an examination at a distance, they could contact the HCC and order an examination, and after that, the person could see their physician at a distance again. They also saw opportunities for contact via RT with the large hospital to a great extent. A participant thought that it would have been even easy to log in via an app or a smartphone:

> Everything ends up on the phone, it seems, and it would have been the easiest thing to do through Messenger or whatever you have for video, it would have been the ultimate, then you do not even have to go anywhere. [ID8; female; aged 53 years]

A theme was interpreted from the 6 categories—participation and relationships are important for good and close care via RT—and was about people experiencing participation and relationships in different ways. When people felt involved and experienced good relationships and reliable technology, a sense of safety and security was created during the RT meeting. When patients felt less involved and the relationship or technology was not satisfactory, a feeling of distance and insecurity provided less good care via RT.

Discussion

Overview

In this study, we investigated the experiences of patients with cardiovascular diseases regarding follow-up meetings with their physician through RT. The inductive analysis showed a common thread throughout the categories that was interpreted as the theme, participation and relationships are important for good and close care via RT. Close care can mean more than just geographical proximity. The importance of relationships can be interpreted as the availability to meet those in care that we already know. A digital tool such as a video screen can also convey a sense of proximity.

Owing to the high demand for PCC, as described in the Introduction section, we further viewed our results considering the 6 categories developed by Sharma et al [37] in an overview of reviews of PCC. However, these components are compiled from various sources and are not mutually exclusive. Therefore, our results may fit in several of the PCC categories. We found that such an approach works well, but certain aspects of PCC in RT are missing. Therefore, we suggest that when RT is introduced, the PCC categories need to be extended with digital aspects in each of the components: establishing a therapeutic relationship, getting to know the person, shared power and responsibility, empowering the person, trust and respect, and communication [37]. To simplify the structure of the discussion, we abbreviate this digital extension as electronic PCC (ePCC) in analogue with eHealth.

Principal Findings

Establishing a Therapeutic Relationship

PCC Partnership

This component is based on a partnership with mutual dependency and responsibility between the person and the professionals, and key factors are open communication, a cohesive team, and professionals who possess knowledge and skills to practice PCC [37].

ePCC Partnership

Our results showed that a therapeutic relationship should be established in person before an RT meeting. This relationship created security, which was confirmed by informants in the category, safety with good relationships, and the subcategory, familiarity with the staff. Some were also skeptical about RT with an unknown person, and a participant claimed, “But had it been a complete stranger then you get a little...then you keep a little distance...” Participants preferred to see a physician they were familiar with and trusted in a video consultation, which was confirmed in another study [20].

Getting to Know the Person

PCC Holistic View

This component emphasizes a holistic view of the person or patient that is more than the illness or disease that the person is diagnosed with [25,26]. It is vital for professionals to seek, understand, and acknowledge the experiences, values, and wishes of the patients and what is relevant to them. Another essential key factor is to get to know the person’s family and their culture to be able to provide care that is adapted to the patient’s need [37].

ePCC Holistic View

In this study, participants felt closeness via the screen and described an experience of personal contact in the category, proximity and distance with technology. Lavoie et al [38] believe that we enter an ethical relationship as soon as we meet a foreign face. The face expresses a meaning, and we must respond to the message of the face. For example, the experience with facial expressions described previously in this category shows that RT may work well for some. However, some other participants missed the body language and felt unsure about how to interpret the therapist’s reactions. The size of the screen and the quality of the sound and image were of great importance for how participants perceived the RT meeting. Some of the participants in this study experienced a personal contact via the screen, and some felt a distance and experienced it as impersonal, which gave them a feeling of insecurity. A feeling of alienation has been described in previous studies [17]. In the study by Shulver et al [39], some participants preferred physical visits because it was more personal, conducting the videoconference alone was isolated, and human contact was important. In our study, for example, a participant wanted a physical meeting but was never consulted. In addition, our study showed that there may be an added value for a physical meeting outside the RT meeting such as social interaction with the taxi driver or doing some shopping.

Therefore, the caregiver needs to know the person’s digital literacy and ability to communicate via ICT and whether the person has any disabilities such as visual, hearing, or cognitive impairments or those that require or which a physical meeting.
**Shared Power and Responsibility**

**PCC Shared Responsibility**

The component, *shared power and responsibility*, indicates that the patient needs to be an active part in their care, and the care delivered should be individualized and based on the person’s own needs, wishes, and values [37].

**ePCC Shared Responsibility**

There was a willingness among the patients in this study to be prepared for the RT meeting, which could be to write down questions in advance and could be seen in the category, *participation in care*, and the subcategory, *being prepared*. However, new ways of meeting could also be demanding for some people; thus, they forgot to ask questions [40], which also was described in the category, *quality of and familiarity with technology facilitating the meeting*, and participants described feelings of not being in control. Some informants felt comfortable because they were using this technology in everyday life, and others felt uncomfortable at the beginning of the meeting; however, later in the conversation, they stated that “...there was nothing strange about that.” The staff are also responsible for ensuring that the patient can use the technology before they organize an RT meeting, and they are also responsible for being able to handle the technical equipment themselves. Informants in this study commented that in some meetings, the physician was uncomfortable with the technology, and in another RT meeting, the person saw just half of the physician’s face.

**Empowering the Person**

**PCC Empowerment**

This component highlights the importance of the individual being active in their own care. The staff needs to provide patients with information, support, and resources that make it possible for them to be able to make their own decisions [37].

**ePCC Empowerment**

In this study some participants felt that the care via RT was equivalent to a physical meeting described in subcategory, *equitable care*, where patients were surprised that the digital meeting gave the same results as a face-to-face meeting, which is also described in other studies [15,20]. Participants in the study by Johansson et al [15] thought that patients would receive the same care regardless of whether they met the specialist physically or via a digital meeting. As the care is experienced as equivalent regardless of the type of visit, web-based or physical, the care becomes more accessible to the people who live in SPRs. This saves both time and money, as they do not need to travel to the HCC or hospital, and this was experienced by the individuals in the category as *good accessibility*. If the person can be in their own context where they feel safe, it is a way of empowering the individual, and the RT meeting can be as effective as a physical meeting [37].

The informants wanted to continue to meet via RT and saw opportunities to develop digital meetings further, where even mild ailments could be treated, something that could be seen in the category, *participation in care*. Another study showed that some participants had a wish for video meetings in the future, to avoid unnecessary trips to the hospital, especially because the informants became old and for other reasons [15]. Allowing patients to become involved and obtain information about how the technique works before the meeting and how this influences the development of care strengthens people’s participation in their own care [37].

**Trust and Respect**

**PCC Personal Needs**

The trust and respect component involves recognition of the person as a unique individual with their own values, preferences, lived experiences, and needs. Health care practitioners need to consider the person’s individualized needs and incorporate them into their care [37].

**ePCC Personal Needs**

In this study, the RT meetings can be equated with a physical meeting with a physician or other health care professional, where confidentiality and personal integrity are important. The informants in this study pointed out the importance of not having other people in the same room or people passing by the room in the category, *cherishing personal integrity*. For the patients who participated in an RT meeting, it was important to trust their physician, and to have a mutual respect to emphasize privacy. Patients in this study highlighted the importance of the room for digital meetings being in such a way that no other people were present except for those who were to attend the meeting, which was also found to be important in another systematic review [16]. New demands are placed on health care staff in the ambition to provide PCC for patients in connection with digital meetings, and technology provides opportunities for patients to influence future care, but there is an increased risk of integrity being violated. On the basis of our results and those of previous studies, it is important to protect personal integrity; there is always the risk of lack of confidentiality owing to a lack of control of the physical rooms where the meetings are conducted, for the patients, physicians, and technical devices used. It is even more important to decide whether it would be an RT meeting because some patients prefer to travel to a physical meeting because of social and practical needs and an added value from the journey.

**Communication**

**PCC Information**

This component about communication between the person, their family, and health professionals is important to discuss and deliver understandable and correct information about the person’s care [37].

**ePCC Information**

Communication is the most salient difference from face-to-face meeting because it is mediated through digital devices in RT. Our results showed that RT affects the person before, during, and after the meeting. This means that the other PCC components need to be taken into consideration when planning for an RT meeting.

*Before the RT meeting, it is important to establish a therapeutic relationship, get to know the person,* and
empower the person as mentioned previously. This concerns the person’s vision, hearing, cognition, and IT literacy. Some participants in this study had hearing problems and difficulty in perceiving what was being communicated, which was experienced as insecurity. People with hearing loss may misunderstand advice or instructions, which could be prevented by using a headset or if the physician, a nurse, or an accompanying person could be present at the meeting and explain what is said [15]. The fact that people with disabilities have access to an interpreter is also a way to improve communication and promote the mutual relationship with the patient, and it is consistent with Swedish law [6]. In this study, a relative supported a person with hearing and memory difficulties—in the category, safety with good relationships.

Communication and information about how the technique works and how the meeting will be conducted are at least as important as communicating about the patient’s illness and health care for an optimal RT meeting. Some participants felt it unusual to communicate via a video screen for the first time, and some were used to it as shown in quality of and familiarity with technology facilitating the meeting, which was also confirmed by Johansson et al [40].

Moreover, in this study, the staff connected the device for some patients and started the meeting, which normalized the meeting and made the patients more comfortable; this is also noted in the study by Currie et al [41]. Patients felt supported when the staff connected the equipment and explained how the digital meeting would be conducted [17,20].

To master care via RT, health care professionals need to acquire knowledge and understanding about how digital technology affects the interaction between people. Technology has an important supporting function in the meeting between health care professionals and patients, but there is also a risk of the technology contributing to frustration and alienation [42].

Finally, it may be noted that an RT meeting may not be suitable for some people, as explained in the category, proximity and distance with technology.

During the RT meeting, it is important to share power and responsibility by ensuring technical quality and control conditions both at the caregiver’s site and at the person’s site, which may affect trust and respect.

For a person to be perceived as more sympathetic and present in a digital meeting, they should look at the camera and not at the face on the screen. The camera should also be placed at the minimum eye level; however, the distance may be less decisive [43].

In this study, it was easy for some informants to communicate in the RT meeting when the sound was good, whereas others experienced a feeling of distance and could not be spontaneous in the conversation; for the latter, it was difficult because they could not see body language and facial expressions, which made them feel distant in proximity and distance with technology.

Technology could also be disruptive as shown in the subcategory, support or disturbed by technology. An informant, for example, saw only half of the face and thought it was embarrassing to tell the physician. Similar incidents were reported by participants from another study when they could not see the computer screen because it was placed incorrectly and did not understand that they could ask for the screen to be placed differently [15]. If the picture was small and placed in a corner, the meeting felt unnatural. The use of technical equipment can be frightening for inexperienced users, and it is important to consider each person’s needs—what information and instructions they need to be able to use the new technology, so that they feel comfortable and safe.

To limit the experience of distance and isolation, the health care service can appoint a person to support the patient during the visit [38]. If the digital meeting is conducted from home, perhaps a relative can support the person or a staff member can support the person if it is from an HCC unit. Of course, no other unauthorized people should be present, as was unfortunately reported by some patients—discussed previously in the component, trust and respect.

Other authors believe that mobile ICT should be used with caution and that health care professionals need to understand older users, and the lack of knowledge about modern technology can affect the person’s attitude toward the new technical solutions [41,44].

After the RT meeting, it is important for the patient to feel safe, with no need to worry about the equipment. In this study, a patient “felt awfully bad...” about the fate of the equipment if it was still turned on—in the subcategory, want more information. As described in the category, participate in care, it is necessary to know the end of your responsibility when the meeting is over.

Strengths and Limitations

Only 8 people participated in this study, which can be a weakness, but there was a diversity of sex and age represented. There were also difficulties in including more patients, as there were few patients who had an RT meeting. However, in the remote regions and SPRs, it is a proportionally good representation because there are few people living in this area. In the last interviews, no new information emerged, which could be a sign of saturation [31]. In this study, none of the participants were younger than 53 years. On a group level, younger people may have higher digital literacy than older people. However, in PCC, group-level properties cannot be generalized to the individual. It is important to assess each person’s abilities using ICT, regardless of age, which is highlighted in our results.

Each person’s experience is unique, and there is no perfect truth to be found, but common patterns and individual differences have been reproduced and described. The results are comparable with those of previous studies with other patient experiences in digital meetings, which also confirms the transferability of the results from this study. Qualitative analysis was used because the purpose was to describe experiences of a phenomenon, a new method for meeting, and a qualitative method is therefore
suitable [32]. The analysis process has been conducted close to the text with a low level of abstraction and often returned to the original material to maintain the holistic perspective, and no text materials have been excluded from the analysis [32,45]. The codes, subcategories, and categories were discussed among the authors throughout the process. A detailed description of the process and the analysis, with quotes from the interviewees, have been explained in the Methods and Results sections for a comprehensive understanding of the patients’ experiences of RT meetings and to increase the trustworthiness of the results. Our interpretation is that the categories cover the data well and can be confirmed by the quotes, thus increasing the credibility of the study.

Implications for Future Studies and Practice

Future studies are needed to identify additional factors that affect telemedicine acceptance, such as human-technology interaction, the organization of the health care system, and social and cultural human factors. Information and education about how digital services work in practice are especially needed for patients and professionals who lack technical skills.

Patients felt that close relatives were a support, and it is important to interview relatives’ experience of RT and the staff’s perceptions. More systematic studies are needed about how people experience a digital meeting and for whom this way of meeting is suitable depending on individual conditions, resources, what disease or diseases and symptoms the person is affected by, and what the cultural context means for the willingness to seek care. For some individuals, the best way to connect to the physician may be from home. In the future, RT, in addition to patients not having to travel long distances and saving time, may also result in an economic benefit for people, communities, and health care systems as care becomes closer and more accessible. Thus, there is also a great need for health economic evaluations of digital meetings in health care. It is an interesting area of research regarding what savings can be made in terms of climate impact and sustainable development when travel is reduced.

Finally, as technical development is exponentially fast and both professionals and older people will probably have better skills, high digital literacy, and more experiences with various platforms, the results of this study may not be repeatable over time; however, other factors may turn out to be as important for optimal RT meetings.

Conclusions

This study has shown that participation and relationships are important for good and close care via RT. To improve the quality of an RT meeting, PCC can be applied but needs to be extended to ePCC, especially the communication component as the most salient difference from a face-to-face meeting.

Before an RT meeting it is crucial to do the following:

1. Establish a therapeutic relationship, get to know the person, and empower the person, preferably at a previous physical meeting.
2. Decide whether the meeting should be held in person or via RT based on the person’s preferences, abilities, and social and practical needs.
3. If conducting an RT meeting, get to know the person’s vision, hearing, and cognitive abilities and digital literacy and take measures, if necessary.
4. Acquire knowledge and understanding of how digital technology works and how to manage it among the health care staff.

During the RT meeting it is vital to do the following:

1. Ensure technical quality and control conditions both at the caregiver’s site and at the person’s site.
2. Ensure that the meeting rooms are designed in a safe and secure manner and that privacy and confidentiality are ensured.
3. Place the camera at the minimum eye level.
4. Look at the camera and not at the face on the screen.

After the RT meeting, it is important to do the following:

1. Ensure that the patient feels safe and that they do not need to worry about the equipment.

RT meetings need to be created with various actors within the care organization based on a person-centered approach, where the patient is a co-creator of good and close care in the future.

Acknowledgments

The authors would like to thank all participating patients who shared their experiences.

Authors' Contributions

CE and AE-L were responsible for designing the study in addition to the interviews conducted by CE. AE-L and CE analyzed the interviews and prepared the manuscript. P-DL contributed with comments and critically revised the text from a pedagogic, juridical, and technical perspectives. All authors commented and approved the final version of the manuscript.

Conflicts of Interest

None declared.

References


Abbreviations

COREQ: Consolidating Criteria for Reporting Qualitative Research
DHR: digital health room
ePCC: electronic person-centered care
HCC: health care center
ICT: information and communication technology
PCC: person-centered care
RT: remote treatment
SPR: sparsely populated region

Edited by A Kushniruk; submitted 30.09.22; peer-reviewed by N Davoody, D Valdes; comments to author 01.01.23; revised version received 12.03.23; accepted 13.04.23; published 01.06.23.

Please cite as:
Ek C, Liljegren PD, Edin-Liljegren A
Patients With Cardiovascular Disease Revisiting Specialist Physicians via Remote Treatment: Interview Study of Experiences
JMIR Hum Factors 2023;10:e43125
URL: https://humanfactors.jmir.org/2023/1/e43125
doi:10.2196/43125
PMID:37261892

©Charlott Ek, Per-Daniel Liljegren, Anette Edin-Liljegren. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 01.06.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Decision Aids for Patients With Head and Neck Cancer: Qualitative Elicitation of Design Recommendations From Patient End Users

Eleah Stringer1,2, RD, MSc; Julian J Lum3,4, PhD; Jonathan Livergant1,5,6, MD; Andre W Kushniruk1, PhD

1School of Health Information Science, University of Victoria, Victoria, BC, Canada
2Nursing and Allied Health Research and Knowledge Translation, BC Cancer, Vancouver, BC, Canada
3Trevor and Joyce Deeley Research Centre, BC Cancer - Victoria, Victoria, BC, Canada
4Department of Biochemistry and Microbiology, University of Victoria, Victoria, BC, Canada
5Radiation Oncology, BC Cancer - Victoria, Victoria, BC, Canada
6Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada

Abstract

Background: Patients with head and neck cancer (HNC) carry a clinically significant symptom burden, have alterations in function (eg, impaired ability to chew, swallow, and talk), and decrease in quality of life. Furthermore, treatment impacts social activities and interactions as patients report reduced sexuality and shoulder the highest rates of depression across cancer types. Patients suffer undue anxiety because they find the treatment incomprehensible, which is partially a function of limited, understandable information. Patients’ perceptions of having obtained adequate information prior to and during treatment are predictive of positive outcomes. Providing patient-centered decision support and utilizing visual images may increase understanding of treatment options and associated risks to improve satisfaction with their decision and consultation, while reducing decisional conflict.

Objective: This study aims to gather requirements from survivors of HNC on the utility of key visual components to be used in the design of an electronic decision aid (eDA) to assist with decision-making on treatment options.

Methods: Informed by a scoping review on eDAs for patients with HNC, screens and visualizations for an eDA were created and then presented to 12 survivors of HNC for feedback on their utility, features, and further requirements. The semistructured interviews were video-recorded and thematically analyzed to inform co-design recommendations.

Results: A total of 9 themes were organized into 2 categories. The first category, eDAs and decision support, included 3 themes: familiarity with DAs, support of concept, and versatility of the prototype. The second category, evaluation of mock-up, contained 6 themes: reaction to the screens and visualizations, favorite features, complexity, preference for customizability, presentation device, and suggestions for improvement.

Conclusions: All participants felt an eDA, used in the presence of their oncologist, would support a more thorough and transparent explanation of treatment or augment the quality of education received. Participants liked the simple design of the mock-ups they were shown but, ultimately, desired customizability to adapt the eDA to their individual information needs. This research highlights the value of user-centered design, rooted in acceptability and utility, in medical health informatics, recognizing cancer survivors as the ultimate knowledge holders. This research highlights the value of incorporating visuals into technology-based innovations to engage all patients in treatment decisions.

(JMIR Hum Factors 2023;10:e43551) doi:10.2196/43551

KEYWORDS

decision support; decision aid; app design; oncology; head and neck cancer; patient information needs; qualitative
Introduction

Head and neck cancer (HNC) is the sixth leading type of cancer by incidence worldwide and diagnosed in approximately 4300 Canadians per year [1]. Based on clinical and pathological manifestation, patients with HNC may be provided an option of surgery, radiation, chemotherapy, immunotherapy, and combinations thereof, each influencing morbidity differently [2]. For patients to participate in their care, they must comprehend their disease and treatment options, consider their own preferences, participate in decision-making to the degree they wish, and make a decision consistent with personal preferences. Patients who direct decisions, even if more than anticipated, fare better on all decision-related outcomes, emphasizing the need for oncologists and surgeons to endorse, facilitate, and support patient participation in treatment decision-making [3].

Some patients with HNC may struggle to take part in decision-making regarding treatment options because of the complexity of the information that needs to be accurately conveyed and understood [4,5]. This can be particularly challenging in the context of a globalized world consisting of cultural differences and varying health literacy levels, increasing the need for enhanced transparent communication of risk and outcomes associated with treatments [5,6]. Furthermore, research on risk literacy in medical decision-making shows that across different cultures, people often struggle to grasp the prerequisite concepts necessary for understanding health-related risk information such as numbers, graphs, and basic medical facts. Errors occur because inappropriate information formats complicate and mislead adaptive decision makers [6].

Decision aids (DAs) “are interventions that support patients by making their decisions explicit, providing information about options and associated benefits/harms, and helping to clarify congruence between decisions and personal values” [7]. DAs complement, rather than replace, counseling from a health care practitioner as an interactional strategy to facilitate patient involvement and contribute to patient concordance [4]. DAs are useful when the best treatment strategy depends on a preference for the benefit-harm trade-off inherent in a particular choice.

A scoping review conducted on electronic DAs (eDAs) for patients with HNC returned 12 relevant articles that discussed 5 different patient eDAs [8]. The scoping review confirmed the value of eDAs in this population supporting “further research and development in this area.” The scoping review did not, however, reveal detailed technical features of existing eDAs. The patient eDA developed by Petersen et al [9] remained available online, so features were viewable, but the other 4 eDAs were not available for viewing. The aim of this study is to respond to the gap in the literature on preferred DA architecture by interviewing survivors of HNC on the utility and potential visual designs for an eDA for patients with HNC to encourage informed and collaborative decision-making. Through phenomenological inquiry, the goal was to answer the following questions: (1) What do survivors of HNC think about the utility of electronic decision-support tools utilizing visuals for patients with HNC? and (2) What suggestions do survivors of HNC have on a potential design for visual features and core components of a prototype eDA?

Results will be used to inform next steps in the development and integration of eDAs in HNC care.

Methods

Prototype Visual Development

A point-of-reference (ie, example screen visualizations) mock-up of a DA was developed for the interviews (Figure 1), containing a graph used in the symptom management clinic at BC Cancer-Victoria (BCC-Vic). A line for surgery is included in the graph, serving as a placeholder only, as surgery is managed by surgeons external to BCC-Vic. The graph was displayed to participants in the interviews. Two symptoms were selected for visual demonstration, oral mucositis (Figure 2; photo credit [10]) and radiation dermatitis (Figure 3; photo credit [11]), noted by a shape on the curve. Hovering the cursor over these shapes triggers a pop-up image of the side effect, increasing in severity from left to right in accordance with the Common Terminology Criteria for Adverse Events (CTCAE) [12] grade. The information used in the design and graphs shown to participants was based on the evidence-based literature on HNC [13-15]. A key for these shapes was not provided, as they were meant to provide a demonstration of the mouse-over effect, where information could be obtained by moving the cursor over areas of the screen.
Figure 1. Severity of side effects graph for a prototype decision aid (displayed to patients during interviews). RT: radiation therapy.

Figure 2. Example image with Common Terminology Criteria for Adverse Events grading and name (mucositis). Credit [10].
Recruitment and Materials
Two phases of recruitment were conducted through convenience and purposive sampling, respectively, with a target of 6-12 participants or until no new knowledge was obtained [16]. Inclusion and exclusion criteria are described in Textbox 1. Upon consent, demographic, diagnosis, and treatment plans were extracted from BCC’s electronic medical record.

Textbox 1. Inclusion and exclusion criteria for the semistructured interview.

**Inclusion criteria**
- Diagnosis of head and neck cancer (including nasopharyngeal) of any staging with completion of treatment (radiation or chemotherapy, with or without surgery)
- Treatment completed within the previous 5 years
- Resides on Vancouver Island or Gulf Islands

**Exclusion criteria**
- Inability to communicate proficiently in English
- Inability to meet via a virtual platform
- Participation discouraged by oncologist, psychiatrist, or other physician

Interviews and Analysis
One-hour long, semistructured interviews, including field notes and reflective questions [17], were conducted following an interview guide. Interview questions were open ended and fell under the categories of introduction, background, experience, e-tools, evaluation, and conclusion. The interview questions were designed to elicit participants’ experience in learning about side effects of treatment along with their impression of the visuals presented to them (in terms of how well they helped explain treatment and side effects; see Multimedia Appendix 1 for the full interview script). For consistency, interviews were conducted by a single researcher (ES), who is most familiar with the subject area and experienced in patient interviews. No other researchers participated in the interviews to remove potential power dynamics that may disempower or affect participant opinions. Participants were introduced to the interviewer including their motivations and background behind this research.

Interviews were held virtually through Zoom (Zoom Video Communications, Inc.), to allow for screen sharing while also ensuring participants safety during COVID-19; all participants attended from their home. Participants were provided a demonstration of the eDA but did not interact with it. Participants were encouraged to ask for further demonstration if needed and were welcomed to use Zoom’s annotate feature. All interviews were video-recorded, transcribed, and analyzed using Taguette software [18,19]. Reflexive thematic analysis was performed on annotated transcripts, field notes, and reflections, where an open and iterative process was applied to coding following the 6-step approach of Braun et al [16]. A set of tagged codes (see Multimedia Appendix 2 for the codes) was used to identify concepts in the transcripts related to the research questions. These codes were used to develop concepts or
“domain summaries at the start of the analytic process.” [16] Upon further engagement and critical reflection, related concepts were then grouped into themes [16]. Coding was completed by the researcher ES who was most familiar with the data and then verified by AWK. Data validation was conducted by sharing preliminary themes with participants for correction, modification, or confirmation.

Ethics Review

Review Ethic Board (REB) approval was granted on January 18, 2021, from the University of Victoria Human REB (University of Victoria Study #BC20-0546) and UBC/BC Cancer Agency REB (H20-02307). All participants provided written informed consent to take part in the study.

Results

Demographics

A total of 12 participants were interviewed: 6 recruited from the Head and Neck Support Group at BCC-Vic and 6 identified by an oncologist. Participant ages ranged from 34 to 76 years, with half (n=6) of the participants diagnosed with stage III cancer and nearly half (5/12, 42%) undergoing surgery prior to radiation. As many as 7/12 (58%) participants were offered adjuvant chemotherapy: 3 accepted cisplatin and the remaining 4 declined chemotherapy; 9/12 (75%) participants accessed the symptom management team on a regular basis, while 10/12 (83%) connected with a counselor through treatment (Table 1).

Table 1. Participant demographics.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Frequency, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>8 (67)</td>
</tr>
<tr>
<td>Female</td>
<td>4 (33)</td>
</tr>
<tr>
<td>Age</td>
<td></td>
</tr>
<tr>
<td>30-40</td>
<td>1 (8)</td>
</tr>
<tr>
<td>41-50</td>
<td>2 (17)</td>
</tr>
<tr>
<td>51-60</td>
<td>3 (25)</td>
</tr>
<tr>
<td>61-70</td>
<td>2 (17)</td>
</tr>
<tr>
<td>71-80</td>
<td>4 (33)</td>
</tr>
<tr>
<td>Cancer staging&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Stage I</td>
<td>2 (17)</td>
</tr>
<tr>
<td>Stage II</td>
<td>4 (33)</td>
</tr>
<tr>
<td>Stage III</td>
<td>6 (50)</td>
</tr>
<tr>
<td>Stage IV</td>
<td></td>
</tr>
<tr>
<td>Treatment type</td>
<td></td>
</tr>
<tr>
<td>Surgery</td>
<td>5 (42)</td>
</tr>
<tr>
<td>Radiation</td>
<td>12 (100)</td>
</tr>
<tr>
<td>Chemotherapy</td>
<td>3 (25)</td>
</tr>
<tr>
<td>Cisplatin</td>
<td>3 (25)</td>
</tr>
<tr>
<td>Cetuximab</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Declined by participant</td>
<td>4 (33)</td>
</tr>
<tr>
<td>Degree of involvement with the symptom management team (general practitioner in oncology and registered nurse)</td>
<td></td>
</tr>
<tr>
<td>Less than once per week</td>
<td>3 (25)</td>
</tr>
<tr>
<td>Weekly</td>
<td>5 (42)</td>
</tr>
<tr>
<td>Greater than once per week</td>
<td>4 (33)</td>
</tr>
<tr>
<td>Degree of involvement with patient and family counseling</td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>2 (17)</td>
</tr>
<tr>
<td>1-3 times through treatment</td>
<td>3 (25)</td>
</tr>
<tr>
<td>Greater than 3 times through treatment</td>
<td>7 (58)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Based on the American Joint Committee on Cancer 8th edition [20].
Thematic Categories
A total of 61 codes were developed and applied, pointing to 17 concepts related to eDAs and the prototype. These 17 concepts were organized into 9 themes, separated into 2 categories (Table 2).

<table>
<thead>
<tr>
<th>Category</th>
<th>Themes</th>
<th>Concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>eDAs(^a) and decision support</td>
<td>• Familiarity with DAs(^b)</td>
<td>• Support of concept</td>
</tr>
<tr>
<td></td>
<td>• Support of concept: usefulness and value of visual aids in explaining treatment and its side effects</td>
<td>• Appreciation for learning with visuals</td>
</tr>
<tr>
<td></td>
<td>• Versatility of prototype</td>
<td>• Communication</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Patients would use the app in different ways</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Design and features</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Emotion, traumatic experience, and resiliency</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Emotional impact of physical changes that can be seen by others</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Trend for a specific symptom to leave a large, lasting impact</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Coping strategies</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Altruism</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Meditation/mindfulness</td>
</tr>
<tr>
<td>Evaluation of mock-up</td>
<td>• Reaction to prototype</td>
<td>• Recommendations</td>
</tr>
<tr>
<td></td>
<td>• Favorite features</td>
<td>• Appreciation for the care team</td>
</tr>
<tr>
<td></td>
<td>• Preference for customizability</td>
<td>• Areas for improvement within the care team</td>
</tr>
<tr>
<td></td>
<td>• Complexity</td>
<td>• List of practical suggestions</td>
</tr>
<tr>
<td></td>
<td>• Presentation type</td>
<td>• Differences in experiences before and during COVID-19</td>
</tr>
<tr>
<td></td>
<td>• Suggestions for improvement</td>
<td>• Value of family and connecting with other patients</td>
</tr>
</tbody>
</table>

\(^a\)eDA: electronic decision aid.
\(^b\)DA: decision aid.

Category 1: eDAs and Decision Support
The term “decision support tools” or “decision aids” was new to 75% (9/12) of participants, while 3/12 (25%) were familiar due to their line of work (Table 3). Regardless of eDA experience, there was strong support for the concept of using visual aids to enhance the explanation of treatment options, their potential side effects, and timeline, demonstrated by SUPPORT OF CONCEPT being the most frequently applied code (see Multimedia Appendix 2).

All participants expressed the usefulness and value of photographs, videos, and graphs. Learning styles appeared to influence the degree of value placed on graphics, with visual learners demonstrating the most enthusiasm. Those whose information needs were adequately met through their oncologist’s verbal explanation still suspected images would reinforce and amplify their message.

The codes USEFUL and VALUE were the most frequently tagged alongside PHOTOS and VIDEOS, often describing statements on how viewing photos of side effects would have clarified and adjusted treatment expectations. Lastly, several participants felt that viewing severe side effects prior to initiating treatment would serve as a catalyst, motivating prophylactic, intensive therapy and self-care strategies to optimize outcomes and quality of life. For example, one participant felt that viewing images of radiation dermatitis prior to treatment would have improved their compliance with skin care guidelines.

An unexpected theme was the potential versatility of eDAs in this area. Participants envisioned the eDA serving not only as an educational tool, but also as a communication tool to use both within and outside clinical encounters. Of the 12 participants, 3 (25%) felt that an eDA would prepare them for oncologist appointments by prompting the creation of informed questions. Participants also described their interest in using an eDA with the interdisciplinary care team, friends, and family. Two participants, for example, described the exhaustion that resulted from the draining task of updating loved ones on their day-to-day well-being.
### Table 3. Representative quotes for thematic category 1: eDAs\(^b\) and decision support.

<table>
<thead>
<tr>
<th>Theme and subthemes</th>
<th>Representative quotes</th>
</tr>
</thead>
</table>
| 1. Familiarity with DAs\(^b\) | • “That’s the whole point of my life is to provide the data for those tools.”  
• “I have [heard of DAs] in terms of like breast cancer screening.” |
| 2. Support of concept: usefulness and value of visual aids in explaining treatment and its side effects | • “I think an app would be awesome!”  
• “And it seems like it’s really prime and the right time to be investing in more technologies, just because they’ve received such an expansion. Now, of course, everyone’s watching health dollars and how much healthcare is costing.”  
• “It’s an exciting initiative, I think it’s great to try to incorporate technology into improving care and treatment.”  
• “I like this, I like what you’re doing here. I like this concept very much. Good for you, you know, getting early into the treatment and looking at side effects. So this is very effective already.” |
| Visual learners | • “I think it would be great for me because I’m a visual learner, but I think it’d be great for other people because, even those who are not visual learners, because I found it, like- I’m a reader, and I found it difficult to read at certain times and also things I read, I read it and then re-read it and I’d be like, ‘What does this even say?’ I couldn’t answer simple questionnaires, I was like, ‘I don’t even know what they’re asking.’”  
• “I think visuals are very helpful, audio and visual you know if you have like a recording, or you have slides on your screen because I tend to be a visual learner. Anyway, and it’s something that you can actually re-read yourself. It’s very concrete. And it’s very well organized.”  
• “I’m a visual learner so I, I really think that would benefit me, so yeah, very interesting.”  
• “I think the use of pictures work for me. I’m a visual person so that would work for me, videos and depending on the topic I think videos are quite useful as well.” |
| Visual to augment the message | • “[The prototype] would have augmented [the oncologist’s] message. So yeah, I think any visually would help.”  
• “I remember they sent me in a room with one of the [radiation] technicians with a sheet of possible side effects. But it was very, again, it was very technical, and kind of, you know, ‘Do this. Don’t do that. You might have this. You can use this cream,’ but I think if there’s been a tool like this, for them to go through that would have been really helpful.”  
• “If I can see pictures, or if I’m given handouts, or if the oncologist speaks more slowly, or shows me pictures. Yeah, that would have been more helpful to me.” |
| Clarifying and adjusting treatment expectations | • “I think an image is worth a thousand words.”  
• “I think seeing a picture like that would have been prepared me more.”  
• “I like the graphic images that come up because then you can see ‘Oh yeah, these, these sores in my mouth or this burning in the neck area okay that’s what it looks like,’ and it’s like, ‘yeah, that’s what I went through.’”  
• “And that really helps clarify things so, yeah, a lot of information.” |
| 3. Versatility of a prototype | • “I’d want to study [the app] and then I’d have questions and then if I went in [to my appointment] and [my oncologist] said, ‘Oh, here’s what to expect,’ and I had already looked at it and could think of my questions.”  
• “Early on, because you just don’t really understand what the process looks like. But if you have seen some of these photos, you might be like, oh okay I should ask questions about that.”  
• “Yeah, definitely. And probably ahead I would want to look at it ahead of time, and then discuss.” |
| Interdisciplinary care team | • “I would probably carry it with me to each medical appointment. Because you know, the first question is, ‘Well, how are you this week?’ So I’d just opening the chart and say, ‘Well I felt like I was here [pointing at chart] even though I should have been here [pointing]. Can you explain why I have not experienced this side effect?’ Yeah I think it’d been perfect.”  
• “How are you feeling today?” because everyone asks you those questions. And now you can just show them. ‘See I’m feeling this [pointing to chart]. This is how it’s changed on the [chart] today.’”  
• “It would be helpful to bring it to the nurses, like, you know, during treatment I didn’t see my oncologist that often. The nurses would be helping me with most of these symptoms and things so they would have access to that as well maybe you could read it and discuss it with them.”  
• “I personally liked it and I think it helps when you’re trying to pass [information] on to somebody else so we use that sort of system, again [in my work] for making decisions.”  
• “There was a hard part of trying to explain symptoms that I had.” |
Sharing with friends/family

- “I think I would really want to share this [eDA] with the people close to me. And so that that would be pretty neat if rather than, you know, people say, ‘How are you doing?’ and you say, ‘Oh I’m doing good, I’m doing good,’ or ‘This is what’s happening;’ it would be neat to be able to share whatever the news is, with being able to show other people a graph of what’s happening.”
- “And then [my support] would have a better idea. If you said, ‘I’m in week four.’ They might not know. Is this the end or just halfway through or...? But this provides some really good context to what you’re going through.”
- “My oncologist was just maybe a really good communicator, but like [using an eDA] didn’t seem necessary. But this would have been really helpful when I’m explaining things to other people, you know, when my parents were like, ‘What’s going to happen?’ Would be nice to have something like this to, to show them like it would help me explain it better.”
- “So I think the person who supports the people, because they’re going to look in your mouth and go, ‘That’s what’s happening.’ I think that’s normal right and maybe they even feel comforted that, that looks right you know.”
- “...encourage people to share it so that the people in their life like, get it because it’s hard to understand, you don’t see a lot of what goes on, it doesn’t show right it’s not like a cast.”

“eDA: electronic decision aid.

bDA: decision aid.

Category 2: Evaluation of Core Visual Components of a Prototype eDA

Initial reactions to the screens and visualizations that were presented to participants during the interviews were positive, with several participants exuding immediate enthusiasm. All participants described their favorite features in a desired prototype eDA. The visualizations presented were most praised for their simplistic design using a combination of pictures, colors, graphs, and text that offered a snapshot of treatment (Table 4).

Although there was unanimous agreement on the usefulness and value of visuals shown, the least consistent responses regarded the complexity of the graph. All participants felt they could read the graph without an explanation; some even read it “instantly.” While some felt the eDA visualizations may be too simplistic, one-third of participants (n=4) expressed a concern that the graph is overly complicated for the public, which includes individuals of varying education levels, ages, and health literacy. One participant expressed concern that the use of technology in medical care may be disconcerting to the elderly population, but suggested that if the difficulty is technology navigation, then the oncologist can operate the device.

Although there was strong support for the basic features of the prototype, every participant provided suggestions for improvement that pointed to their desire for customized information. Some mentioned features to consider customizing including treatment type (option to view a single treatment only), radiation dose (total radiation grays or number of fractions), filtering by side effect, adjustable timelines, results stratified by sex, detail within photos, option to display information source/reference, and associated statistics.

There were mixed opinions on the appropriateness of medically explicit images that become increasingly gruesome as the CTCAE grading score advances. Every participant agreed to view these images, but several appreciated the invitation to skip. Those in favor of including vivid detail argued that it is an accurate and honest display of treatment realities, while those hesitant to include graphic photographs were concerned that it would instill fear and possibly dissuade the pursuit of treatment. Two participants suggested offering the option of viewing an artistic sketch with less detail instead of a photograph. Another participant suggested emphasizing that symptoms are manageable with guidance from the health care team, using this as a segue to discuss supportive care measures such as medications and self-care techniques. Furthermore, they suggested it is “really important” to make clear that these strategies will “not cure it but will help them.”

There were mixed responses on which type of device (eg, desktop/laptop, tablet, phone) a prototype should be designed for, discussing the advantages and limitations of each device. Those with strong opinions based their response on how they envisioned themselves using an eDA. For example, one participant suggested a mobile phone interface for the benefit of portability, as they envisioned using it on public transport. A different participant preferred a computer to view details on a larger screen. Lastly, another participant thought a tablet to be the most practical for clinic use as it could be passed between the patient and health care provider while using a screen large enough for comfortable viewing.

SUGGESTIONS FOR IMPROVEMENT was the third most frequently applied code. Most recommendations fit into the following 7 categories:

- Additional information to add: Participants offered creative suggestions on information and features to add to the prototype, summarized in Table 5
- Legend: As previously described, 2 shapes dotted the graph representing 2 different side effects. These shapes were used for demonstration purposes only and would not be part of the final prototype, but participants found this confusing and were concerned that symbols would complicate the graph. They liked the idea of selecting the side effects of interest from a list on the side.
- Links to additional resources: Two participants highlighted the opportunity to link to supportive care resources, such...
as Inspire Health, BCC Patient and Family Counseling, and meditation exercises.

- **List of side effects:** Several participants recommended expanding the list of side effects to encompass rare side effects, such as fatigue, headaches, hair loss, neuropathy, tinnitus, vision problems, and dermatological changes.

- **Interactive tracking:** There was great enthusiasm for building out interactive tracking options. Some ideas were options to track diet intake, speech language pathologist–prescribed exercise routine, mental health, and symptoms such as pain and nausea.

- **Add videos:** Several participants suggested adding videos, with 1 participant particularly adamant on the value video could add, including “video clips of people who were able to speak to their experience.”

- **Late side effect/life after treatment:** Several participants requested additional information on life after treatment.
### Table 4. Representative quotes for thematic category 2: evaluation of mock-up.

<table>
<thead>
<tr>
<th>Theme and subthemes</th>
<th>Representative quotes</th>
</tr>
</thead>
</table>
| 4. Reaction to prototype | ```
- “I can tell you right off the bat, my reaction is this is fantastic. Oh, that’s really great! That was good.”
- “I like how clean and simple this is”
- “I like it. Great idea, and then that ended up just seeing that graph. Okay. It clicked. I, again, I think it’s awesome and again that’s just the way I do things the way I make decisions like that.”
- “Awesome! The graph to me, it’s wonderful.”
- “Gold star!”
``` |
| 5. Favorite features | ```
- “This is brilliant, especially what you just described, if you could choose your side effects, or you could choose the side effects you’re concerned about like, you’re like, ‘I’m tired.’ Okay I got that I’m tired, that’s fine. I don’t need to learn more about that tired – I get that, but for the ones that are maybe, ‘Oh skin rash burn doesn’t sound good,’ maybe, you know, like the ones that seemed more serious or more concerning to be able to click on those because I was just thinking let’s say you had a shake for every possible side effects.”
- “Oh golly, well since you kind of explain the symbols and the curves and the free categories and time and all that. This is fantastic for my use because when you hover in the pictures come up then I can put it all together. And I think it’s very helpful because it’s important to know if you’re going to look in your mouth. And you see something that doesn’t look like your mouth used to look like. Then you can refer back to here and say, ‘Oh, I get it.’ Okay, did this happened to me during week for during week six during. Not at all. And yeah, I think this is very cool because it uses color is just a basic graph. It’s very simplistic but yet you know and then it hovers and you get photographs which you know every photograph almost looks like me. I could relate to all of that.”
- “I mean when I first looked at this without those the pop-up pictures, it’s good but then you just see that the severity goes up, but you can’t visualize it or you can’t internalize what that means. So I think being able to click on those pictures is great because that can give you an idea.”
- “[The prototype’s] essential because just looking at numbers, and like numbers in a row doesn’t really mean much. But looking at pictures, charts and graphs and however you want to display that information, particularly with color is a far better way of communicating content.”
- “You’re very smart to have been able to control when you show the photographs here, because you know a lot of presentations I have all the photographs and you just click on each photo and yet, I think that might frighten too many people away. Wo it’s good then that only when you hover is when the related photograph shows up.”
- “I like the chemo and radiation like to have them where they are parallel.”
- “I love the list of potential side effects”
``` |
| 6. Complexity | ```
- “...dead simple, you know, the two axes the severity over time. Yeah, I don’t know, I’m biased by background and math and physics and things like that but you know that, you know, It makes it makes perfect sense”
- “This is pretty self-explanatory. You know, if I’m going through radiation, I’m just going to look at the blue line. It’s so simple.”
- “I think it’s good for adults, because it simplifies, and it just minimize this confusion when you break things down into words, pictures, colors, sound.”
- “Seriously, the KISS [Keep It Simple Stupid] theory. This is perfect, because the more fancy, you’re going to be bombarded with questions and complaints and who knows what. So I think the simpler the better.”
- “If you’re not well educated something like this could leave somebody awful confused.”
- “You have to assume there’s patients who have no education and they would not be able to make sense of this.”
- “I would worry about especially...about older patients with, you know, some might have cognitive or dementia. They would go, ‘What is this this? Some kind of mathematical thing? I don’t understand this.’ and they wouldn’t use it.”
- “It might be to assume the level of education, or sophistication of the patients. And if you want to make it accessible for everyone, then, then it might need a different might need to show it in a different way.”
- “I would just worry about the graph...especially for patients who have no scientific training whatsoever.”
- “It’s fine but, you know, not everybody relates to graphs.”
- “I like that idea of it but I would say, why would you need a graph if you’re going to have a visual? Why doesn’t one of you just have someone come on and say, ‘This is a picture that could happen you could look like this after three weeks, four weeks, you could look like this after five weeks,’ rather than a graph where someone is trying to figure out what the graph is.”
``` |
| 7. Preference for customizability | ```
- “If there’s any ability to customize that would be a really useful feature.”
- “The simple graph is good, especially if there’s any way to customize it to be a little bit more like you check [the side effects you want to see].”
```
<table>
<thead>
<tr>
<th>Theme and subthemes</th>
<th>Representative quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Support of graphic detail</td>
<td>• “I like the graphic images that come up because then you can see, ‘Oh yeah, these sores in my mouth or this burning in the neck area, okay, that’s what it looks like’ and it’s like ‘yeah that’s what I went through.’”</td>
</tr>
<tr>
<td>Opposing detail</td>
<td>• “[The oncologist] using this has to support that this is the worst-case scenario.”</td>
</tr>
<tr>
<td></td>
<td>• “I guess as long as a person knows these are possible side effects and not necessarily your degree of severity. Yeah, I guess that’s right. Yeah, no I think that’s good.”</td>
</tr>
<tr>
<td></td>
<td>• “I like the idea of that, but the only thing that you’ve given us is some fear factor a little bit. If you look in there and say, ‘Geez I don’t know if I should go through this because look what’s going on.’ Because I thought that’s what’s going to happen to me, but it didn’t. I was a bit sunburnt looking but nothing like what you showed there- nothing. And I didn’t get the most sores. So, as long as, ‘This is the worst case scenario.’”</td>
</tr>
<tr>
<td></td>
<td>• “I don’t mind those pictures. In fact, like I said when I look at that and go, ‘Well that didn’t happen to me. Wow—that’s bad! Like I’m glad I didn’t have that,’ but what if it didn’t happen here? What if yours was worse than that or, like, I think, pictures are really important, but maybe if someone had an option of either a drawing or a picture as opposed to a picture or no picture.”</td>
</tr>
<tr>
<td></td>
<td>• “So for people that are a little bit nervous or don’t know what the word is- they feel they get the upset stomach thing from seeing [graphic photos], if they had a picture, not a photo, I mean like an artist drawing, so that they have an idea...It doesn’t show the same thing- it’s not the same thing! But if someone can at least see a neck and then some dots on it, but they really don’t want to see someone’s neck burn because that’s going to make them sick, just to have that option...so everyone’s involved.”</td>
</tr>
</tbody>
</table>

8. Presentation type

|  | • “I don’t use an iPad, and I do wonder how much is generational. I just use my phone and my laptop for work. But, like my parents, and I think that a lot of this cancer is in an older age group, that they use their iPads all the time. So I think for me it would be phone and laptop but I recognize that might be a generational thing. I don’t have an iPad. I think most people have phones, right, if they have any of those devices.” |
|  | • “For me personally, I have a preference to use it on the computer screen computer, like the regular desktop...I don’t like to carry a lot of electronics around with me, so I don’t always carry my phone.” |
|  | • “I’m 52, I’m going to choose a laptop anyway, because it’s bigger. I wear reading glasses. So I like to see it big.” |
|  | • “I’m looking at this on the phone right now and it makes perfect sense” |
|  | • “The bigger the screen, the more accessible.” |
|  | • “You can have something that’s able to be used in all three devices.” |

9. Suggestions for improvement

|  | • “Find ways to visually simplify it a little bit.” |
|  | • “Maybe you want to put a comments area” |
|  | • “I was thinking what’d be great on the app, if there could be a little pop up.” |

Legend

|  | • “I just see those five [symbols on the prototype] and what the relationship they are. So, I really agree with not having too many because it was just a road of shapes you’d be like, you know, it’s too much, but to be able to choose it or to pick a few of the [side effects]” |
|  | • “I might do something different with that, like, I don’t know if triangles and circles are the best way” |

Links to additional resources

|  | • “I was thinking it would be great on the app, if there could be a little pop up: ‘Don’t forget Inspire Health’ or whatever other resources, you would like to add to it.” |

List of side effects

|  | • “I love the list of potential side effects but as I say I would include others if you know other like eyes, definitely.” |
|  | • “I love your suggestion of adding in the interactive tracking possibilities that would be very helpful” |

Interactive tracking options

|  | • “I like tracking things and graphs. If it didn’t exist, I would do something on my own that made it interactive, so that I would track every day.” |

Add videos

|  | • “I think video would be the way to go. It’s for a person my age. You read. What do you retain? Keep reading. I think if I could see somebody, the exercises they were performing. I think that would be much more informative and easier to grasp.” |
|  | • “I think it’s an excellent idea to have a video that people. The doctor, maybe can go say I’m just going to leave the room for a few minutes, watch this video for five minutes or whatever it is. And I think that’s a great idea and have the patient check that and, and, or have the doctor sitting there for doctrine to target wait to wait another five minutes, but I think is a great idea or I’ll send a video home with the patient.” |
|  | • “I’m a visual person so that would work for me, videos and depending on the, on the topic I think videos are quite useful as well.” |
Theme and subthemes

Late side effects/life after treatment

• "I think that would be another thing is warning about late side effects."

• "What I'm looking for is, you know, some sort of post treatment timeline that says at one year. Most of our people are here at two years, most of our people are here these are the some of the side effects they're experiencing. Oh and look you know 40% of these people are having cognitive issues of some kind, just like they just get mentally tired more quickly than they used to do."

<table>
<thead>
<tr>
<th>Table 5. Recommended additions to the prototype.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recommendation</td>
</tr>
<tr>
<td>Include a clarifying disclaimer statement</td>
</tr>
<tr>
<td>Add photos of the radiation bed, chemo room, and other commonly used clinical areas</td>
</tr>
</tbody>
</table>
| Extend the timeline | • "The only thing I would add is maybe like a line vertical bar for vertical line for one year out another one for 10 years ago."
• "My thinking is that those dotted gray lines could move like I don’t mean you should move them I mean they could"
| Integrate self-care and coping strategies | • "I do think that some kind of an app [could be linked]. Even that could allow you to practice some techniques for coping with the actual treatment, like I said, like a breathing where you could watch a short video on.” |
| Add an area for feedback | • "I was gonna say maybe you want to put a comments area but none that’s another good idea because adults always have comments, but that’s about it.” |
| Add an option to view statistics | • "I'm interested in things like well, of all your patients what percent just did RT, what percentage did chemo and RT what you know of all those combinations surgery chemo. How did it break out, like, you know, what are the what are the treatment paths that people have been on. What are the result pass for those treatment combinations and what are the side effects down the road. I'm a data person so I wanna, I want to make sure there’s the data behind that” |
| Advertise other relevant research and clinical trials | • "I like your suggestion to have seen if there’s a way to link it with that other study that’s currently going on where you’re inputting things.” |

Discussion

Principal Findings

This study contributes to the literature on the utility and patient-recommended requirements and design of eDAs. Participant feedback was overwhelmingly supportive of using visuals to support explanations of treatment and potential side effects. All participants could imagine themselves using an eDA as an education and communication tool and agreed to contribute to the next phase in the full development of eDA, further demonstrating their endorsement.

Use and Value of Visual Aids

Literature demonstrates that patients prefer images to demonstrate benefit and harm trade-offs in health as they are perceived as easier to understand [21]. Verbally relayed numerical facts were least effective in encouraging a specific health decision while graphical information was the most preferred, demonstrating that “consideration should be given to developing visual aids to support shared clinical decision making” [21]. This is consistent with our findings: every participant agreed that the use of images alongside verbal explanations would complement or augment an oncologist’s message. Furthermore, “using transparent information formats enhances risk comprehension, communication, and recall and helps people make better decisions about their health” [22].

Malleability of the eDA

Participants were interested in adding interactive options; most were intrigued by symptom tracking to auto-populate a personalized curve on the graph. Here lies an opportunity to incorporate patient-reported outcome measures, standardized instruments used to capture patients’ perceptions of their health status, functional status, or health-related quality of life [23]. Integrating patient-reported outcome measures within the eDA would support the gold standard of patient self-administration without interviewer interpretation [24]. Furthermore, literature
demonstrates the benefit of real-time symptom tracking during cancer care, improving health outcomes and communication with health care providers [25,26].

When considering participants’ desired features alongside cumulative feedback, customizability is required to suit the range of learning styles, needs, and preferences. A potential solution to varying information needs may include suggestions from Table 5 and then allowing users to filter and sort the viewable data. For example, participants unanimously supported the idea of selecting side effects of interest from an extensive list of potential side effects. Taken further, graphics could be shown as real-life photos or sketched images to better meet participants comfort level with medically explicit images. Embedding customizability may also address the discourse regarding the suitability of the graph. The mixed responses align with the literature on graph literacy, which “reveal[s] that people, regardless of their numeracy skills, differ substantially in their ability to understand graphically presented quantitative information about health” [22]. Furthermore, Nayak et al [27] tested graphical interpretation of a visual dashboard on patients with prostate cancer and despite 78% of participants having a college education, variation remained in graph literacy results.

**Prognosis Versus Quality of Life**

Our visualizations designed and presented to participants did not include information on prognosis or survival outcomes to simplify the inception eDA. The information presented was designed to be used in the presence of an oncologist who could personalize the message and discuss prognosis on a case-by-case basis. Previous research, however, demonstrated that comprehension of medical information on survival and cancer treatment options is equal or better than when patients are shown the same information with the addition of mortality statistics [28]. In a viewpoint paper on his personal experience with advanced stage tongue cancer, Kushniruk [29] argues for “patients to be more informed about choices and statistics, including the meaning of survival curves in relation to different treatment options.” Although the desire for incorporation of information about prognosis in DAs was not a prominent theme in our findings, the importance of survival cannot be overlooked and should be considered for inclusion in future iterations.

**Limitations**

Selection bias was introduced through the convenient sampling method. Recruitment initially targeted patients at BCC-Vic’s Monthly Head and Neck Support Group, comprising those closely matching the inclusion criteria. This group, however, may more likely represent survivors who experienced challenges through treatment, thus introducing a source of bias. To help offset this, an equal number of participants (n=6) were recruited through purposive sampling to diversify the demographics. Because of the nonprobability sampling methods, the results cannot be generalized to a wider population. Future work should consider including the perspective of those without cancer (ie, cancer naïve) and survivors along the cancer trajectory (ie, never diagnosed through to long-term survivorship). Furthermore, race was not included in the demographics and future work should include diverse racial representation, including historically underrepresented groups.

There are limitations to the mock-up design method used for presenting design ideas and visualizations that did not adhere to a formal development process, such as the Ottawa Patient Decision Aid Development eTraining or the International Patient Decision Aid Standards Collaborations [30]. It was instead based on the research teams’ phenomenological and personal experiences with HNC; graphs used with patients at BCC-Vic; informal, preliminary feedback from patients; and results of the scoping review. The designs and visualizations were sufficient for this study purpose of applying a user-centered design [31] to this innovation with plans to build through the co-design methodologies of Kushniruk and Nohr [32] and Kushniruk and Patel [33].

Repeat interviews to improve data richness were not conducted as these interviews were intended to be preliminary, setting the stage for the next phase of the project that will include in-depth workshops with participants. Additionally, rich data were collected from the single interviews and additional interviews were not yet deemed necessary.

Lastly, due to the specificity of the target population, results cannot be generalized to other tumor types. Further requirements analysis and testing will be required prior to expanding to other populations.

**Conclusions**

This research highlights the value of incorporating visuals into technology-based innovations to support patient decision-making in oncology care. All participants felt an eDA, used with their oncologist, would support an enhanced and transparent explanation of treatment and augment the quality of the consult. Participants liked the simple design of the prototype visualization but desired customizability to adapt the eDA to their individual information needs. This research highlights the value of user-centered design, rooted in acceptability and utility, in medical health informatics, recognizing cancer survivors as the ultimate knowledge holders.

The next steps include applying the co-design methodologies of Nayak et al [27] and Kushniruk [29] to develop a fully functioning eDA to be tested with patients with cancer, incorporating visualizations and the feedback described in this paper. This will consider implications to clinical workflow, including physician’s time, and improve accessibility by ensuring visualization is amendable to printing for those without digital access. Usability testing of the eDA will be conducted in a clinical setting at BCC-Vic and the University of Victoria through a newly funded follow-up study (ie, the Head and Neck Cancer Application for Patients and their Partners [HANC APP Study]).
Acknowledgments

We thank the participants who shared their personal experiences through treatment and provided valuable insight into this project. This work was graciously supported by The BC SUPPORT Unit-Vancouver Island Centre.

Authors’ Contributions

ES developed the study design, mock-ups, conducted the interviews, analyzed the results, and wrote the manuscript. JJL provided study oversight and edited the manuscript. JL provided medical oversight and assisted in recruitment. AWK contributed to the study design, provided oversight, verified analysis and study results, and edited the manuscript.

Conflicts of Interest

ES received funding for this project from the BC SUPPORT UNIT-Vancouver Island Centre, and receives research funding from Michael Smith Health Research BC and the Lotte & John Hecht Memorial Foundation. AWK receives research funding from the Natural Sciences and Engineering Research Council of Canada (NSERC) and is the Editor-in-Chief of JMIR Human Factors at the time of this publication. JLL receives research funding from the Canadian Institute for Health Research (CIHR). JL has nothing to declare.

Multimedia Appendix 1

Interview questions.
[DOCX File, 17 KB - humanfactors_v10i1e43551_app1.docx]

Multimedia Appendix 2

Frequency of tagged codes.
[DOCX File, 19 KB - humanfactors_v10i1e43551_app2.docx]

References


18. Taguette, the free and opensource qualitative data analysis tool. Taguette. URL: https://www.taguette.org/ [accessed 2021-05-06]


Abbreviations

BCC-Vic: BC Cancer-Victoria
CTCAE: Common Terminology Criteria for Adverse Events
Elected by Y Gong; submitted 14.10.22; peer-reviewed by B Chaudhry, I Soares-Pinto; comments to author 11.12.22; revised version received 28.02.23; accepted 19.04.23; published 05.06.23.

Please cite as:
Stringer E, Lum JJ, Livergant J, Kushniruk AW
Decision Aids for Patients With Head and Neck Cancer: Qualitative Elicitation of Design Recommendations From Patient End Users
JMIR Hum Factors 2023;10:e43551
URL: https://humanfactors.jmir.org/2023/1/e43551
doi:10.2196/43551
PMID:37276012

©Eleah Stringer, Julian J Lum, Jonathan Livergant, Andre W Kushniruk. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 05.06.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original work on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Developing Implementation Strategies to Support the Uptake of a Risk Tool to Aid Physicians in the Clinical Management of Patients With Syncope: Systematic Theoretical and User-Centered Design Approach

Abstract

Background: The Canadian Syncope Risk Score (CSRS) was developed to improve syncope management in emergency department settings. Evidence-based tools often fail to have the intended impact because of suboptimal uptake or poor implementation.

Objective: In this paper, we aimed to describe the process of developing evidence-based implementation strategies to support the deployment and use of the CSRS in real-world emergency department settings to improve syncope management among physicians.

Methods: We followed a systematic approach for intervention development, including identifying who needs to do what differently, identifying the barriers and enablers to be addressed, and identifying the intervention components and modes of delivery to overcome the identified barriers. We used the Behaviour Change Wheel to guide the selection of implementation strategies. We engaged CSRS end users (ie, emergency medicine physicians) in a user-centered design approach to generate and refine strategies. This was achieved over a series of 3 qualitative user-centered design workshops lasting 90 minutes each with 3 groups of emergency medicine physicians.

Results: A total of 14 physicians participated in the workshops. The themes were organized according to the following intervention development steps: theme 1—identifying and refining barriers and theme 2—identifying the intervention components and modes of delivery. Theme 2 was subdivided into two subthemes: (1) generating high-level strategies and developing strategies prototypes and (2) refining and testing strategies. The main strategies identified to overcome barriers included education in the format of meetings, videos, journal clubs, and posters (to address uncertainty around when and how to apply the CSRS); the development of a web-based calculator and integration into the electronic medical record (to address uncertainty in how to apply the CSRS); a local champion (to address the lack of team buy-in); and the dissemination of evidence summaries and feedback through email communications (to address a lack of evidence about impact).
Conclusions: The ability of the CSRS to effectively improve patient safety and syncope management relies on broad buy-in and uptake across physicians. To ensure that the CSRS is well positioned for impact, a comprehensive suite of strategies was identified to address known barriers.

(JMIR Hum Factors 2023;10:e44089) doi:10.2196/44089

KEYWORDS
emergency medicine; physicians; qualitative research; risk management; syncope; user-centered design

Introduction

Background

Syncope is a prevalent and high-cost problem in emergency departments (EDs) defined as a sudden transient loss of consciousness followed by spontaneous complete recovery [1-4]. It accounted for 15,476,451 ED visits in the United States from 2005 to 2015 [5]. In Canada, there are approximately 160,000 ED visits annually that translate to direct hospital costs of approximately CAD $130 million (approximately US $101 million) [6,7]. Although the cause is often benign, approximately 10% of patients will have serious underlying conditions identified within 30 days [8,9]. Up to half of these conditions will be identified after the initial decision to discharge or admit the patient from the ED, emphasizing the need for supports to improve risk stratification and decision-making in the ED [1,8-10].

The Canadian Syncope Risk Score (CSRS) is a validated risk stratification tool to optimize the accuracy of ED decisions and inform evidence-based clinical actions [9,11]. The tool encompasses the calculation of a risk score from which the following evidence-based practice recommendations for ED disposition are derived: immediate discharge of low-risk patients without subjecting them to further unnecessary testing, consideration for a short course of hospitalization for high-risk patients, and discharge of medium-risk patients with cardiac monitoring and clear information regarding the serious outcome risk. Similar to many decision aids, the uptake of the CSRS is likely to prevent unnecessary hospitalization and improve outcomes for those with underlying conditions through the implementation of a standardized and evidence-based approach to ED syncope management. However, the existence of the CSRS is only part of the solution—physicians must see the value and use the tool in practice to realize the expected benefits.

Health services research shows that many evidence-based practices or interventions fail to demonstrate the anticipated impact as they were not properly implemented, precluding them from achieving a positive impact on clinical practice and patient outcomes [12,13]. Implementation failure might be because of an insufficient understanding of the context in which the implementation process occurs [14,15] or an absence of barriers identification influencing the use of such interventions [16]. The field of implementation science provides systematic approaches and strategies to address the research to practice gap by systematically assessing likely barriers to uptake and identifying implementation strategies to target these barriers [13,17-19].

Objective

The objective of this developmental work was to describe the process of developing a set of implementation strategies to support the use of the CSRS in real-world ED settings to improve syncope management among physicians. This systematic approach not only identifies which strategies are the most appropriate for the target context but also how they should be implemented and operationalized to mitigate the risk of suboptimal implementation or poor uptake among target users [20].

Methods

We followed the process for intervention development outlined by French et al [21], including the first three steps: (1) identifying who needs to do what differently, (2) identifying the barriers and enablers to be addressed, and (3) identifying the intervention components and modes of delivery to overcome the identified barriers.

Study Design

Steps 1 and 2: Identifying Who Needs to Do What Differently and the Barriers and Enablers to Be Addressed

We previously identified that emergency medicine physicians need to change their approach to the assessment of syncope, which would influence their subsequent management among patients. The initial qualitative work identified barriers among 41 physicians across 12 Canadian ED sites to both CSRS use and the adoption of its evidence-based practice recommendations [22]. The most salient barriers identified were workflow issues, concerns about continuity of care, the lack of confidence in the CSRS, and the lack of knowledge and skills around how to interpret and apply the CSRS-related criteria for various patient profiles [22]. The most reported enablers were as follows: legitimacy in the decision rule (CSRS), the evidence of safety and benefit to send the patient home, cardiologists and emergency medicine physicians buy-in, and adequate time with the patient.

The understanding of these barriers and enablers was refined and contextualized with other groups of emergency medicine physicians who participated in workshops as part of a user-centered design (UCD) approach, which is presented in step 3. Essentially, the list of barriers was presented to physicians, and they were asked to react out loud to the following questions: Does this list look complete? Are there other barriers you would like to bring in our attention? and What do you think?
Step 3: Identifying the Intervention Components and Modes of Delivery to Overcome the Identified Barriers

This step was informed by the Behaviour Change Wheel [23], in which we systematically mapped the barriers identified in qualitative work [22] to theoretical determinants, as presented in the Theoretical Domains Framework (TDF) [24] and the Capability, Opportunity, Motivation, and Behavior (COM-B) model [23]. We went through the qualitative findings and looked at physicians-related barriers and facilitators (e.g., the lack of knowledge around the eligibility criteria of using CSRS) and matched them with the corresponding theoretical determinant (“knowledge” [TDF] and “capability” [COM-B]). TDF offers a comprehensive lens to look at cognitive, affective, social, and environmental factors [24] that can influence CSRS uptake, whereas COM-B allows for broader categories of determinants.

Once this granular and systematic theoretical understanding of determinants has been completed, we linked them to evidence-based behavior change techniques (BCTs) [25,26], which are the active ingredients or components of an intervention. This step allows to select the most likely techniques to produce the desired change [23,26,27]. From the literature [23,25,26,28], we identified the most effective BCTs that can address each of the behavioral determinants [23,27]. We defined each BCT, how it addressed determinants, and how we could operationalize them. We assessed the feasibility of using these BCTs through peer debriefing (GR, LD, and Marlena Dang Nguyen) and by using the Acceptability, Practicability, Effectiveness, Affordability, Side-effects, Equity criteria [28]. An excerpt of the process of identifying the intervention components and modes of delivery to overcome the identified barriers is presented in Multimedia Appendix 1 [23,25-30].

Selecting those intervention components (also named “implementation strategies”) was the theoretical groundwork that fed the subsequent steps of our developmental process.

We used a UCD approach to validate and refine our understanding of the barriers and enablers and to identify modes of delivery and operationalization. We engaged emergency medicine physicians who were the end users of the CSRS under a collaborative, participatory, and cocreative lens to pursue the parallel goals of maximizing usability in the context of those targeted by the implementation endeavor and tailoring strategies to users’ local contexts while retaining the core components responsible for their effectiveness [18]. This was achieved over a series of 3 qualitative UCD workshops lasting 90 minutes each with 3 groups of emergency medicine physicians. Data collection was performed on the web through synchronous interactions using the Zoom videoconferencing platform (Zoom Video Communications Inc). The workshop-related processes and content are summarized in detail in Textbox 1. Examples of probing questions in the workshop facilitation guide are presented in Textbox 2. GR facilitated the workshops. She did not know the participants before the study. The facilitator (GR) encouraged a “think-aloud” approach to provide insight into participants’ thought processes and gather feedback on which implementation strategies might be useful and why (or why not). Specifically, participants were asked to share their reactions, sentiments, and thought processes in real time. The goal was not to achieve data saturation.

We followed the COREQ (Consolidated Criteria for Reporting Qualitative Studies) [31] to report the qualitative process (Multimedia Appendix 2).
Textbox 1. Workshops-related processes and content.

**Workshop 1**
- Material sent: study information, link to complete web-based survey, and 2 scientific papers that support the development and validation of the Canadian Syncope Risk Score (CSRS [9,11]).
- Objectives: (1) solicit feedback on previously identified barriers to uptake the CSRS, (2) rank barriers in terms of priority for attention with the Zoom polling function, and (3) discuss and brainstorm which strategies might effectively address the barriers and improve the uptake of the CSRS.
- Analysis: review notes, review audio recording to summarize perceptions and key insights, debrief with team.
- Outcome: create a mock poster in response to participant feedback for discussion at workshop 2.

**Workshop 2**
- Preparatory work: view existing educational videos (n=4) about how CSRS was developed and validated and what are the underlying practice-based recommendations.
- Material sent: study information, link to complete web-based survey, link to access the educational videos, and 2 scientific papers that support the development and validation of the CSRS [9,11].
- Objectives: (1) solicit feedback on previously identified barriers to uptake the CSRS (workshop 1), (2) solicit feedback on previously identified strategies (theoretical work), and (3) define parameters of operationalization of the strategies (eg, mode of delivery, materials, and content).
- Analysis: review notes, review audio recording to summarize perceptions and key insights, debrief with team, participants’ comments on the workshop summary.
- Outcome: refine the mock poster.

**Workshop 3**
- Preparatory work: view mock poster.
- Material sent: study information, link to complete web-based survey, mock poster, and 2 scientific papers that support the development and validation of the CSRS [9,11].
- Objectives: (1) refine parameters of operationalized implementation strategies identified and (2) discuss the usability and usefulness of the strategies.
- Analysis: review notes, review audio recording to summarize perceptions and key insights, debrief with team, participants’ comments on the workshop summary.
- Outcomes: draw a list of the most salient strategies with operationalized parameters, summarize the implementation strategies, and share this summary with clinical research team that has expertise in syncope management.
Textbox 2. Excerpt of probe questions used in the workshop facilitation guide.

### Workshop 1
- Reviewing the barriers (participant feedback)
  - Does this make sense?
  - Does this resonate?
  - Does this list look complete?
  - Is there anything else you’d like to recommend?
  - Are there other barriers you’d like to bring in our attention?
  - What do you think?
  - Can you talk out loud for me?
  - What do you mean by that?
  - Can you give me an example?

- Co-designing the strategies
  - What are your initial reactions regarding those strategies?
  - Which strategies would encourage you to use the Canadian Syncope Risk Score (CSRS)? What would be helpful for you to use it?
  - What support would you need to implement the tool?
  - Follow-up question: do cardiologists and internists need to be using the CSRS as a decision-making tool in order for physicians to accept it? Or is it simply that cardiologists and internists must accept and buy into the CSRS recommendations?

### Workshop 2
- What type of educational strategy would make sense for you? (eg, educational face-to-face meeting, educational web-based video)
- What are your preferences regarding the mode of delivery of the educational strategy?
- What worked well in the past (when implementing a new rule in your practice)?
- What is the best way to build awareness of best practices across physician colleagues? (ie, WHO helps build awareness, HOW, and WHEN)

**Educational videos**
- What did you like about it? What did you don’t like about it?
- What was effective or not in these videos? (eg, must be shorter, it’s comprehensive)
- What was the right length, duration?

**Prototype posters**
- How useful or not are these QR codes?
- Where should we display this poster?
- In what extent do you think this poster would be effective to act as a prompt and a reminder for you to use CSRS?

### Workshop 3
- Summary of evidence behind the recommendations
  - How useful (or not) are they?
  - How would you like to get access to them? Where should we make them accessible or available?

**Web-based calculator**
- How did you find the web-based calculator?
- Where could you imagine yourself using the calculator?
- At what point during your workflow would you imagine using something like this?
- Where are you when that happens? Are you at a computer? Are you at a desk? Somewhere different depending on what patient you saw? Is there a consistent place?
- How would you like to get access to it?
Ethical Considerations

The qualitative and theoretical developmental work was formally reviewed by the institutional authorities at the Women’s College Hospital and was deemed exempt from a research ethics board approval.

Recruitment

Participant recruitment was facilitated by the main developer of the CSRS (VT) through email communication with key informants at 3 ED sites in Ontario, Canada. Individuals who expressed an interest in participating contacted the project lead (GR) and were provided with a study information sheet via email. In each round, we aimed to recruit between 5 and 7 physicians who had different knowledge levels regarding the CSRS (ie, had heard about it and had basic or good knowledge). We used a combination of convenience and purposive sampling to ensure diversity in sex, hospital site (ie, urban academic and nonacademic), and primary language for clinical care (English or French) to capture various perspectives.

Data Analysis

The workshops were audio recorded and the project lead (GR) listened to the complete recordings to partially transcribe key parts of the conversation to shed light on the barriers and facilitators of using the CSRS, as well as the strategies that would be helpful to improve its use. The project lead (GR) and the research assistant (KW) present during the workshops coproduced a preliminary summary of the findings. We performed a qualitative content data analysis according to the predefined objectives for each workshop by using a deductive and inductive process. We used deductive framework coding with broad categories by applying barriers previously identified as deductive codes as well as theory-based strategies. We inductively coded new strategies and operationalization parameters suggested by the participants throughout the conversation. The research team then reviewed the themes and discussed content changes. A summary of the findings and proposed changes was emailed to the participants after each workshop to solicit feedback; validate the emerging insights and operationalization parameters; and seek clarification, if needed (member checking) [32,33].

Results

Participant Characteristics

A total of 14 physicians participated across the 3 workshops and all were aware of the CSRS before participation. The participant characteristics are summarized in Table 1.

The following two themes were identified from the participants’ perspectives: (1) identifying and refining barriers and (2) identifying the intervention components and modes of delivery. The quotes supporting these findings are presented in Table 2.

Table 1. Participant demographics.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Workshop 1 (n=5)</th>
<th>Workshop 2 (n=4)</th>
<th>Workshop 3 (n=5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex ratio (female:male)</td>
<td>2:3</td>
<td>1:3</td>
<td>2:3</td>
</tr>
<tr>
<td>Years practicing medicine, mean (SD; range)</td>
<td>8.4 (2.7; 4-11)</td>
<td>11.9 (8.53; 4-19.5)</td>
<td>7 (2.62; 3.5-10)</td>
</tr>
<tr>
<td>Site (ED(^a)), n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English academic hospital</td>
<td>2 (40)</td>
<td>2 (50)</td>
<td>3 (60)</td>
</tr>
<tr>
<td>French academic hospital</td>
<td>2 (40)</td>
<td>1 (25)</td>
<td>1 (20)</td>
</tr>
<tr>
<td>English nonacademic hospital</td>
<td>1 (20)</td>
<td>1 (25)</td>
<td>1 (20)</td>
</tr>
</tbody>
</table>

\(^a\)ED: emergency department.
Table 2. Themes supported by participants’ quotes.

<table>
<thead>
<tr>
<th>Theme 1: identifying and refining barriers</th>
<th>Quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discomfort of using the CSRS—a feeling of hesitance</td>
<td>“We reviewed this in journal club at [our hospital] and the biggest thing that came up and common to clinical decision rule is the clinical Gestalt at the end […] Physicians felt hesitant using tool arguing it is telling me if it is vasovagal or cardiac syncope, seems counterintuitive to make decision in order to use tool to tell you something you already know.” [Workshop 2, Participant 2, Female]</td>
</tr>
<tr>
<td>Lack of collective buy-in</td>
<td>“For emergency doctors you are worried to accept the responsibility of discharging people at home. I think there needs to be something coming from cardiologists to say ‘yes, that is acceptable to discharge someone home’; you need to know there is a timely follow-up; it is hard to adopt rule if we don’t feel it is accepted widely by specialists as well. So, we feel supported to safely use it.” [Workshop 1, Participant 1, Female]</td>
</tr>
</tbody>
</table>

Theme 2: identifying the intervention components and modes of delivery to overcome the identified barriers

Subtheme 2.1: generating high-level strategies and developing strategies prototypes

Targeting broad audience when implementing strategies | “Grand rounds, presentations during physicians’ meetings, posters, study sheets all over the place. research assistants remind you to use it; seeing publications, seeing it on social media, pretty embedded in our group; importance of multiple strategies.” [Workshop 1, Participant 4, Male] |

Different formats for educational strategies | “I think case-based rounds is great. I think to get hospital buy-in I’m thinking smaller community hospitals. I think having combined rounds with cardiology, medicine, and emergency to go over the score and how to apply it, and what are monitoring implications—I think that’s helpful as a group. That way the discussion happens with all the key players and the barriers to implementing this. As opposed to presenting it in silos really when a patient comes all these people are important, so combining a strategy could be helpful.” [Workshop 2, Participant 2, Female] |

Poster display at ED | “Flow diagram of what you would do with each category: I think a lot of people in emergency medicine like ‘if this, then that,’ to know which way to flow. That can help to take some of the thought process out if it as long as it is standardized across colleagues/specialists. We want to be practicing along with our colleagues and specialists, so having consensus with colleagues to follow the diagram with appropriate clinical practice and applying the rule appropriately, I think would help too.” [Workshop 1, Participant 4, Male] |

Local champion | “After the six months check-up can be within the department if you have that champion, is the one that can do that link up. The first six months I think will give you enough information, does that local champion can be the one and that links back with the research team and see what is at that point, having someone locally I think is significantly better to get like off the cuff comments and things like that and how they wish it was changed, applied or supports, I think it has better chance of getting quality feedback and regular feedback.” [Workshop 3, Participant 3, Male] |

Subtheme 2.2: refining and testing strategies

Poster | “I can refer to that poster, maybe give me a little bit of credibility if I’m advocating for an admission where I’m getting pushed back.” [Workshop 3, Participant 2, Male] |

Web-based calculator | “Thinking back to other scores, or decision rules that are on calculator…It does bug me sometimes when I’m not able to access like a summary of why that’s the recommendation or why that’s the rule but again having an optional because if you already know it you don’t need to come up every time if you forget or you want to know about the medium risk what exactly are the details having the option to go easily access from the rule would be nice.” [Workshop 3, Participant 5, Female] |

Feedback | This quote speaks to quality indicators that would be of interest: “I think for me anecdotal feedback is really nice.” [Workshop 3, Participant 5, Female] |

Theme 1: Identifying and Refining Barriers

Workshop participants validated the following barriers identified in previous study: discomfort using the CSRS, the lack of confidence, the lack of knowledge and skills, and uncertainty around interpretation. Throughout the workshops, physicians highlighted additional barriers, including struggling with how to apply the CSRS recommendations, the inapplicability of the CSRS for some patient clinical presentations, the lack of the CSRS buy-in from the broader medical team (ie, cardiologists and internists), the lack of evidence about the effectiveness of the CSRS tool, and its practice-based recommendations on patient outcomes. The lack of collective buy-in is an important
barrier to CSRS use, as physicians have described the importance of all team members. The mapping exercise of linking barriers to theoretical determinants is presented in Multimedia Appendix 1, along with examples on how BCTs (eg, credible source, pros and cons, and instruction on how to perform the behavior) can be operationalized.

Theme 2: Identifying the Intervention Components and Modes of Delivery

Subtheme 2.1: Generating High-Level Strategies and Developing Strategies Prototypes

Participants described that effectively addressing the identified barriers required multiple strategies deployed using various dissemination channels. The need to target a broad audience (ie, emergency medicine physicians, cardiologists, internists, head of department, and nursing staff) with consistent exposure over time (ie, repeat messaging) was emphasized. Participants also highlighted the need to leverage existing structures, including integration of the CSRS into the electronic medical record (EMR), discussing the CSRS at standing educational meetings, and displaying a poster in the workplace environment. There was an agreement that educational meetings could be used to promote general awareness of the CSRS and to encourage a nuanced discussion about its application. However, a range of opinions were expressed on the best format for those educational meetings (eg, combined grand rounds and case-based discussions in small groups). Participants felt that holding combined grand rounds with emergency medicine physicians and specialists (ie, cardiologists and internists) would be a coordinated strategy to address multiple barriers simultaneously. How these educational events are promoted is important to stimulate interest and excitement, including highlighting the credibility of the speaker. Having a journal club with emergency medicine physicians and specialists to review evidence around the CSRS would be useful. Podcasts can be another interesting channel to disseminate knowledge around the CSRS. Participants shared the example of Emergency Medicine Reviews and Perspectives [34], a perceived trusted web-based resource, which is a monthly emergency medicine audio series encompassing continuing medical education. Displaying a poster in the ED was suggested as a helpful visual cue, with participants describing the usefulness of the computed tomography head rule poster [35,36] as an example. Participants also highlighted the need for a local champion that could play multiple roles to model the application and use of the CSRS, to influence uptake among colleagues, speak in educational meetings, to facilitate connection between the clinical and research teams, monitor the implementation process over time, and provide in-person or written feedback.

On the basis of these insights, the research team created 2 prototype posters (Multimedia Appendix 3) and prepared a summary of evidence. The 2 posters encompassed similar content with different displays. In poster 1, information around “For whom the CSRS must be applied” and “When to use CSRS” was highlighted. Poster 2 focused on the proposed course of treatment before the application of the CSRS. In both prototype posters, the CSRS was illustrated along with the 3 risk levels and their proposed practice-based recommendations.

In response to participants’ feedback, QR codes were added to the posters: (1) how to use the CSRS, (2) recommendation evidence, and (3) web-based calculator. We prepared a brief summary of evidence to support each practice recommendation for low-risk, medium-risk, and high-risk patients with the intention that physicians feel more confident to use the CSRS and apply the subsequent practice-based recommendations.

The research team also leverages the following existing strategies: educational videos displaying evidence of how the CSRS has been developed and validated and the way to use it, web-based calculators, and email communications used to prompt physicians to use the CSRS and to provide them with positive feedback. The email communication was drawn from a previous pilot study on remote cardiac monitoring as an example to communicate positive feedback on patient impact (ie, an example where home monitoring detected a patient arrhythmia) as well as messaging to remind physicians to use the CSRS.

Subtheme 2.2: Refining and Testing Strategies

Participants discussed the perceived usability, usefulness, and operationalization of the following strategies: educational videos, the poster, the web-based calculator, the summary of evidence, the local champion, and the email communication.

Participants found that the components and features (eg, written summaries, questions, graphics, videos, and links to scientific papers) of educational videos were perceived as useful, and the content was perceived as clear, concise, relevant, and credible. The duration of videos was reasonable if viewed out of the workplace but was too long to be viewed during a shift. It was suggested that a 5-minute video that includes the main information would be an ideal length and would facilitate wider dissemination of the CSRS.

When reviewing the 2 poster prototypes, participants suggested the need to simplify the posters, separating the explicative notes (ie, additional information) from the care pathway, and move those notes as footnotes using a different font (eg, smaller fonts for footnotes) to make the content easier to read. They found poster 2 usable, that is, easy to follow, simple, and appealing. They would use it as a reminder and as a prompt to apply the CSRS, which would be helpful especially at the early stages of the CSRS implementation process in the ED. They would also refer their colleagues to this poster, which is seen as a way of giving credibility to their ED syncope management course of treatment. However, participants identified the following barriers to using such a poster: the risk of poster fatigue, lack of space to display it in their clinical settings, and lack of skills in using QR codes. They suggested to keep only one QR code in the poster, that is, the one related to the web-based calculator.

All participants tested the CSRS web-based calculator [37]. They suggested ways of improving the usability of the web-based calculator: (1) reviewing the wording of some criteria to avoid misleading interpretations, (2) adding a “not drawn” response option to this question “elevated troponin level,” (3) adding access to evidence, and (4) adding access to practice-based recommendations for low-, medium-, and high-risk patients (ie, what to do with the risk score). All
participants intended to use the web-based calculator but for different purposes: use in practice and as an educational tool for medical students. Some would use it only if it is integrated into the EMR and will not use it if it is part of a mobile app. Participants would find it useful to discuss in length the summary of evidence in grand rounds or in another type of educational meeting as an initial evidence uptake. Obtaining easy access to evidence was considered important; tying evidence to a web-based calculator and to the EMR would be one way to improve its access.

Finally, email communication with feedback would be useful for emergency medicine physicians to convince them to use the CSRS. Participants had different opinions on how and by whom feedback could be delivered, such as through educational outreach, one-to-one discussion with local champion, and email communication. They would like to receive feedback from the research team (especially the CSRS developers) and a cardiologist within their hospital.

**Discussion**

**Principal Findings**

The parameters of these strategies are outlined in Table 3. These strategies will be further developed and deployed as part of a nested process evaluation for a stepped wedge cluster trial.

This work was built on a comprehensive and systematic intervention development process anchored in 1 previous qualitative study and in theoretical mapping of linking theoretical determinants with evidence-based strategies. Furthermore, the contribution of this work is to have involved physicians at different stages to gain insight about the perceived barriers and to test strategies in their context. Key barriers included uncertainty about when and how to apply the CSRS recommendations, the lack of resources (eg, cardiac monitors), the lack of buy-in from the broader medical team, discomfort (hesitancy) using the CSRS, and the lack of evidence about the impact on patient outcomes. Surprisingly, no reference on workload or time constraint was brought up, as is often found in other studies [38-41]. Our findings suggest that physician capability should be a central target of implementation supports, specifically the capability to interpret CSRS-based criteria and apply them across a range of clinical presentations.
Table 3. Summary of the strategies and their parameters over the 3 workshops.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Required content</th>
<th>Mode of delivery</th>
<th>Delivery source</th>
<th>Target audience</th>
<th>Target outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Educational meetings (eg, grand rounds) and videos</td>
<td>Nuances, barriers, and pitfalls when using the CSRS; evidence underlying CSRS and recommendations; cost and resources; how to deal with ultra-low-risk criteria and troponin; what to do with risk score</td>
<td>Web-based In-person</td>
<td>CSRS experts, cardiology, and general medicine physicians</td>
<td>All locations where CSRS will be applied</td>
<td>Improve knowledge of and comfort in using the CSRS.</td>
</tr>
<tr>
<td>Web-based calculator</td>
<td>How to deal with troponin criterion; what to do with the risk score</td>
<td>Web2, mobile app, and EMR</td>
<td>Electronic content</td>
<td>CSRS users</td>
<td>Improve CSRS integration into workflow</td>
</tr>
<tr>
<td>CSRS integration into the EMR</td>
<td>Interpretation of the risk score; what to do with the risk score</td>
<td>EMR</td>
<td>Electronic content</td>
<td>CSRS users</td>
<td>Improve CSRS integration into workflow</td>
</tr>
<tr>
<td>Local champion</td>
<td>Roles: speaker, monitor the implementation process, adapt strategies; provide support and feedback. Attributes: strong and positive leadership skills, know how to apply CSRS and recommendations.</td>
<td>In-person</td>
<td>Local emergency medicine physicians and cardiologists (each site)</td>
<td>CSRS users</td>
<td>Improve collective buy-in</td>
</tr>
<tr>
<td>Poster</td>
<td>Care pathway, how to deal with troponin criterion</td>
<td>Paper QR codes</td>
<td>Paper-based</td>
<td>Diverse stakeholders</td>
<td>Improve collective buy-in</td>
</tr>
<tr>
<td>Dissemination of evidence summary</td>
<td>Impact of CSRS practice-based recommendations on patient outcomes. Research papers—CSRS development and validation</td>
<td>On the internet In-person Journal clubs</td>
<td>Electronic content CSRS experts</td>
<td>Diverse stakeholders</td>
<td>Improve knowledge</td>
</tr>
<tr>
<td>Feedback</td>
<td>CSRS impacts on providers’ practice; numbers of cardiac monitor referrals and of arrhythmias detected</td>
<td>In-person and written</td>
<td>Champions</td>
<td>CSRS users</td>
<td>Improve skills and adoption of behavior (CSRS uptake)</td>
</tr>
<tr>
<td>Prompts</td>
<td>Invitation to use CSRS, image with arrhythmia detected (feedback)</td>
<td>Email communication CSRS experts Champions</td>
<td>CSRS experts CSRS users</td>
<td></td>
<td>Improve social opportunity</td>
</tr>
</tbody>
</table>

aCSRS: Canadian Syncope Risk Score.
bEmergency medicine physicians, family physicians working at emergency department, any consultants who are asked for high-risk patients, cardiologists, internists, nurses (including nurse practitioners), and support from head of department.
cMDCalc [37].
dEMR: electronic medical record.
eAll emergency medicine physicians and residents.

Comparison With Prior Work

Training is an evidence-based and frequently used strategy to build physician capability by increasing their skills [23]. This can be operationalized through a variety of mechanisms, including seminars, interactive workshops, and teaching programs such as simulation and training sessions [42,43]. In our study, participants largely referred to educational meetings (eg, combined grand rounds inclusive of all relevant specialties) and educational videos. Skill building can be supplemented by creating increased opportunities to use the CSRS, including integration into the EMR, engaging local champions, displaying posters, and sending email communications to encourage use [40,41,44]. This suite of strategies is commonly used in the ED setting [44] and has demonstrated effectiveness in promoting guideline-adherent care [45].

Achieving collective buy-in across multiple specialty groups (ie, emergency medicine, cardiology, and internal medicine) was highlighted as an essential condition for successful uptake of the CSRS. Although the importance of this broader support is well documented [41,46], including in the area of risk stratification [41], the importance of designing implementation
strategies targeting this broader audience (ie, an audience beyond the immediate end user) has been unexplored. Although ED syncope care primarily rests on the shoulders of the emergency physician, support from experts in cardiology, internal medicine, and hospitals is needed for care of those with suspected or identified serious conditions and further inpatient or outpatient investigations. Physicians rely on their colleagues and professional networks as a unique source of tacit knowledge that serve to either validate initial reasoning or offer alternative approaches [47]. This presents an opportunity to influence uptake through existing channels of social influence that extend beyond the primary setting of interest (in this case, the ED). Strategies may benefit from alignment with the underlying factors that influence patterns of collaboration, including perceived reputational value, experiential information (including personal relationships and visibility), professional identity, and self-awareness of competence [48]. In addition, strategies that target components such as champion or opinion leader, social support, and credible source would be promising ingredients to consider [27].

Similar to the study by Bravo et al [49], our findings highlight a tension between user preferences and scientific evidence and the critical role of triangulating user input in addition to scientific evidence to leverage both sources of knowledge in the design of implementation strategies. Specifically, physician participants identified education in the form of grand rounds as a strategy to address areas of uncertainty and to improve collective awareness of the tool within the interdisciplinary team. Although education effectively increases knowledge and awareness, it is training that effectively builds or strengthens skills [23]. On the basis of physicians’ perceived barrier of not knowing how to apply CSRS, education alone might not be sufficient because it has to do with developing abilities to apply the tool among various patients. In such cases, training would be more suitable. Simply put, each strategy has its own function and mechanisms, allowing it to overcome barriers to using the CSRS and strengthen the facilitators.

Limitations

We used a combination of purposive and convenience sampling to recruit emergency medicine physicians working at 3 different hospitals; therefore, the results may not reflect the experiences of physicians working at other sites. Furthermore, all participants were aware of the CSRS before their participation in the workshops, with most being employed at the same hospital where the CSRS has been piloted. Future studies should explore whether the resulting strategies align with and effectively address the barriers experienced by those who are unaware of the CSRS. Finally, although a comprehensive list of potential BCTs was developed (Multimedia Appendix 1), only a subset of these were prioritized for discussion in the workshops because of feasibility and time constraints. A more comprehensive discussion would have yielded additional strategies to address the identified barriers, and future work should assess whether barriers persist that might be amenable to strategies that were not thoroughly considered. For example, we could operationalize the identification and the preparation of local champions more extensively, and we could target which skills would need to be addressed in a training and how we should impart them (eg, simulation and small-group workshop with demonstration on how to use the CSRS with different patients’ clinical presentations). The provision of performance feedback on the accurate use of the CSRS for risk-stratifying patients by experts (eg, cardiologists and CSRS developers) could also be an avenue to consider.

Conclusions

The ability of the CSRS to effectively improve patient safety and ED syncope management relies on broad buy-in and uptake by physicians. To ensure that the CSRS is well positioned for impact, we identified and developed a comprehensive suite of implementation strategies, including posters, educational meetings (grand rounds), educational videos (with a training component on how to apply the CSRS among various patients), the integration of the CSRS into the EMR, and a web-based calculator to calculate the risk score. These strategies will be evaluated to understand whether and how they are being implemented in practice and whether they are effective in addressing the identified barriers with the objective of improving syncope management in EDs. The next phase of work involves an embedded process evaluation that will provide insight into whether and how this UCD systematic development approach facilitates the ability to effectively target preidentified barriers, physician engagement with the implementation strategies, and broader uptake of the CSRS.

Acknowledgments

The authors are deeply indebted to all emergency physicians at the study sites who participated in that work. The authors would like to acknowledge Dr Maria Keller, Dr Christopher Fabian, Dr Bikalpa Khatiwada, and Natasha Hudek. The authors are grateful to Marlena Dang Nguyen for her input during the early stages of this project. This work was supported by the Cardiac Arrhythmia Network of Canada (CANet; IMPL-002). The opinions, results, and conclusions reported in this paper are those of the authors and are independent from the funding sources. During the study, VT’s research was supported by the Canadian Institutes of Health Research. VT holds a salary award from the Physicians’ Services Incorporated (PSI) Foundation’s PSI-50 Mid-Career Clinical Research Award and a Tier-1 Clinical Research Chair in Cardiovascular Emergencies through the Faculty of Medicine, University of Ottawa.

Conflicts of Interest

VT received an honorarium as a consultant for the project Practical Approaches to Care in Emergency Syncope funded by the National Institute of Health. All other authors declare no other conflicts of interest.
Multimedia Appendix 1
Excerpt of the process of identifying the intervention components and modes of delivery to overcome the identified barriers.

[PDF File (Adobe PDF File), 145 KB - humanfactors_v10i1e44089_app1.pdf ]

Multimedia Appendix 2
COREQ (Consolidated Criteria for Reporting Qualitative Studies) checklist.

[PDF File (Adobe PDF File), 127 KB - humanfactors_v10i1e44089_app2.pdf ]

Multimedia Appendix 3
Prototype posters.

[DOCX File, 428 KB - humanfactors_v10i1e44089_app3.docx ]

References


34. EM: RAP | Emergency medicine reviews and perspectives. URL: https://www.emrap.org/ [accessed 2023-05-23]


Abbreviations

BCT: behavior change technique
COM-B: Capability, Opportunity, Motivation, and Behavior
COREQ: Consolidated Criteria for Reporting Qualitative Studies
CSRS: Canadian Syncope Risk Score
ED: emergency department
EMR: electronic medical record
TDF: Theoretical Domains Framework
UCD: user-centered design

©Geneviève Rouleau, Venkatesh Thirugasambandamoorthy, Kelly Wu, Bahareh Ghaedi, Phuong Anh Nguyen, Laura Desveaux. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 13.06.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors,
is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Factors Reducing the Use of a Persuasive mHealth App and How to Mitigate Them: Thematic Analysis

Markku Kekkonen¹, BA, MSc; Eveliina Korkiakangas², PhD; Jaana Laitinen², PhD; Harri Oinas-Kukkonen¹, PhD

¹Oulu Advanced Research on Service and Information Systems, Faculty of Information Technology and Electrical Engineering, University of Oulu, Oulu, Finland
²Finnish Institute of Occupational Health, Oulu, Finland

Corresponding Author:
Markku Kekkonen, BA, MSc
Oulu Advanced Research on Service and Information Systems
Faculty of Information Technology and Electrical Engineering
University of Oulu
Pentti Kaiteran katu 1
Oulu, 90014
Finland
Phone: 358 029 448000
Email: markku.kekkonen@oulu.fi

Abstract

Background: Studies on which persuasive features may work for different users in health contexts are rare. The participants in this study were microentrepreneurs. We built a persuasive mobile app to help them to recover from work. Representatives of this target group tend to be very busy due to work, which was reflected in their use of the app during the randomized controlled trial intervention. Microentrepreneurs also often have dual roles; they are professionals in their line of work as well as entrepreneurs managing their own business, which may add to their workload.

Objective: This study aimed to present users’ views on the factors that hinder their use of the mobile health app that we developed and how these factors could be mitigated.

Methods: We interviewed 59 users and conducted both data-driven and theory-driven analyses on the interviews.

Results: Factors reducing app use could be divided into 3 categories: use context (problem domain–related issues, eg, the lack of time due to work), user context (user-related issues, eg, concurrent use of other apps), and technology context (technology-related issues, eg, bugs and usability). Due to the nature of the participants’ entrepreneurship, which often interferes with personal life, it became clear that designs targeting similar target groups should avoid steep learning curves and should be easy (quick) to use.

Conclusions: Personalized tunneling—guiding the user through a system via personalized solutions—could help similar target groups with similar issues better engage with and keep using health apps because of the easy learning curve. When developing health apps for interventions, background theories should not be interpreted too strictly. Applying theory in practice may require rethinking approaches for adaptation as technology has evolved rapidly and continues to evolve.

Trial Registration: ClinicalTrials.gov NCT03648593; https://clinicaltrials.gov/ct2/show/NCT03648593

(JMIR Hum Factors 2023;10:e40579) doi:10.2196/40579

KEYWORDS
mobile phone; mobile health; mHealth; Persuasive Systems Design; behavior change; thematic analysis; microentrepreneurs; randomized controlled trial

Introduction

Overview
Health care can be improved by cost-effective solutions with the help of modern health information technologies [1]. In particular, the development of mobile health (mHealth) apps can provide cost-efficient health interventions for a wide range of users, although user preferences may vary considerably [2]. Designing health apps for diverse target groups may seem to be an insurmountable challenge, as stakeholders may have different views on what is important. However, these views are not necessarily mutually exclusive [3]. Even when designing
health apps with special attention to a target group, it is likely that some software features will not be used as much as anticipated by the designers [4].

Engaging users to continue to use health apps is challenging. Persuasive technologies could lessen the challenge, especially if the characteristics of different users are addressed [5]. The persuasive features in digital health interventions supporting users can increase user adherence [6]. According to Fogg [7], persuasion is “an attempt to change attitudes or behaviors or both (without using coercion or deception).”

Designers can use persuasive technologies to motivate people to change their health behavior toward a preferred behavior [8]. Therefore, the use of persuasive technologies to support health behavior change could be beneficial. However, although studies on persuasion extend back to at least 2000 years, persuasion is still not fully understood [8]. Human psychology is complex, and designers may experience challenges when designing persuasive systems [8]. Despite software designers’ best efforts and use of persuasive technologies, getting users to stay active remains a challenge, especially because there are numerous health apps available.

Although research has been conducted on adherence and engagement with digital health apps and interventions [9,10], studies on factors that reduce the use of mHealth apps from users’ perspectives are rare, especially users’ views on how to mitigate these factors.

Designers often add features to an “implementation wish list,” but such features must be justified. Therefore, designers should know which features are persuasive for which target groups. However, this can be challenging if information on the target group is scarce.

Although there are studies on which features or persuasive categories may work for general users in a health context [11] or which persuasive features have been used in specific types of health apps [12], it is more difficult to determine what works for specific groups. Thus, more research is needed on the persuasive features for a variety of target groups.

This study aimed to increase the knowledge on the factors that hinder or reduce the use of persuasive mHealth apps and how these factors can be mitigated. To achieve this, we conducted a data-driven thematic analysis based on interviews (N=59) conducted with users of an mHealth app. In addition, to understand how to avoid pitfalls, we conducted a theory-driven thematic analysis of the interviews. The novelty of this study lies in the persuasion event analysis (data driven) regarding factors reducing the use of the app and the persuasive software feature analysis (theory driven) on ways to mitigate or even improve these factors.

The interviews were conducted as part of an 8-week randomized controlled trial that aimed to help microentrepreneurs recover from work and job strain. The mHealth app for the intervention trial was developed in collaboration with a multidisciplinary research consortium. The Persuasive Systems Design (PSD) [13] model was used as the framework for designing the persuasive technology features within the app. Self-determination theory (SDT) [14] was used as the theoretical background for behavior change, and the transtheoretical model (TTM) [15] was adopted for “Stages of Change”–driven goal setting within the app.

**Research Question**

To gain more knowledge on the topic, we wanted to understand why some users stopped using the app. We also wanted to learn how user engagement could be increased for similar target groups and persuasive mHealth apps.

Therefore, the following two research questions guided this paper:

1. Research question 1: What were the factors hindering or reducing the use of the app?
2. Research question 2: How can the persuasive side of the mHealth app be improved using PSD considering the aforementioned factors?

**Background**

**Microentrepreneurs as the Target Group**

In EU countries in 2014, small and medium-sized companies (SMEs) accounted for 99.8% of all enterprises (in the nonfinancial sector), employing approximately 90 million people [16]. The threshold for defining SMEs in the European Union is up to 250 employees and ≤€10 million (US $11 million) in financial turnover, with smaller firms usually having fewer than 50 employees and ≤€10 million (US $11 million) in turnover [17].

In 2014, about 93% of the SMEs in EU countries were microenterprises [16], which are small companies with ≤10 employees and ≤€2 million (US $2.1 million) in financial turnover [17]. Therefore, in the European Union, microenterprises and microentrepreneurs are vital for national economies. Moreover, in 2016, from 70% to 95% of all firms in all countries were microenterprises, with a large share of those being enterprises with no employees, thus running solely by the microentrepreneurs themselves [18].

Entrepreneurship involves many factors that can cause high workloads, and there is an obvious need to promote work recovery. However, there have been few interventions targeting work recovery in microenterprises [19]. According to Voltmer et al [20], the health of an entrepreneur influences the development of a successful enterprise.

Entrepreneurs are at an increased risk of overexertion [20] because of high responsibilities and demands at work, stress, excessive working hours, fatigue, and sleeping problems [21]. Entrepreneurs also have difficulties balancing work and leisure time [22-26]. Thus, they might benefit from interventions to cope with these professional demands and stress as well as promote healthy behavior patterns [20].

Small businesses are a suitable target group for health promotion [21], but tailored, simple, and low-cost actions are required [27]. Effective recovery from work requires healthy lifestyles [23], including sufficient physical activity, healthy dietary habits [28-31], and stress and time management [23]. Planning beforehand, controlling overtime, having work flexibility, having
social contacts, and exercising regularly are all strategies that can help entrepreneurs maintain good health [23].

**Underlying Theories for the Developed System**

The app used SDT [14] as the theoretical background for users’ behavior change process, thus allowing users to navigate within the system relatively freely. This approach gave the users the freedom to choose any and all content material, tasks, or tools within the app or to choose none. The app also provided relevant and nonjudgmental feedback for the users.

Although users’ self-determination was strongly emphasized, there were minor limitations on user actions within the app owing to development requirements. Before gaining access to the health problem domains, the users had to proceed through 52 baseline questions about their current health behavior, although the questions could be left unanswered. Similarly, in the beginning of each health problem domain module, the users had to proceed through content-specific introductory material once. However, this could also be skipped by pressing the “forward” or “home” buttons.

TTM includes 6 stages of change [15]; however, to avoid complicated goal setting structures for the app, we used an adaptation of TTM. Thus, each module contained 3 goal setting categories based on TTM: think and observe (contemplation and preparation), act and do (action), and maintenance (maintenance). Precontemplation was excluded from the app, as people in that stage would not be ready to proceed toward change and thus could not be engaged. Termination was also excluded because people in that final stage would have no need for the app. Each TTM-based goal setting category contained interactive tasks in all health problem domain modules. After choosing a health domain, users could also choose which stage they wanted, and they were not assigned to any specific goal-setting category by the system. Regarding the first 2 categories, the tasks could be completed either in minutes or within a day or 2.

The tasks in the Maintenance category were supposed to be completed over a longer period, for example, within 10 days. Reminders in the form of push notifications were sent to users who had chosen tasks that required a longer time to complete. Figure 1 shows an edited screenshot of the app (textual content originally in Finnish but translated for this paper).

TTM has been criticized as inappropriate for some behavior change interventions [32]; however, in this case, we feel that the adaptation provided a clear and easy way for users to follow their situations and progress as they worked toward their personal behavioral goals.

Figure 1. The user has triggered a longer task, which will require a few days to complete, depending on when the user wants to complete it.

**PSD Model**

PSD [13] is a model for persuasive software design with design principles for persuasive system functionalities and content. The PSD model offers postulates for describing and evaluating persuasive systems and ways to analyze the persuasion context. There are 4 categories of persuasive principles in the PSD model: primary task support (eg, rehearsal), dialogue support (eg, praise), system credibility support (eg, authority), and social support (eg, normative influence) [13].

The use context, user context, and technology context are the key factors in analyzing persuasive events. The use context includes features or factors arising from the problem domain, such as health behavior. The user context refers to people’s individual differences, including their interests, needs, goals, motivations, lifestyles, and other cultural factors. Regarding the technology context, the strengths, weaknesses, risks, and opportunities of different platforms, apps, and features should be considered [13].
Postulates

According to the first postulate of the PSD model, information technology is not neutral but rather is “always on,” and thus persuasion can be an ongoing process instead of a single act [13]. Therefore, persuasion and persuasive systems require active participation (using the system) from the users, but the system also has to be there for the users for persuasion to happen.

The second postulate of the PSD model emphasizes commitment for cognitive consistency [13]. On one hand, it means that persuasive systems should support and facilitate commitments. On the other hand, users may become committed to performing the target behavior by the support provided by the persuasive system, which naturally means that they should also use the system to achieve this. In terms of SDT, it could be thought that the users should “know” and perform the right actions to achieve their goals, and persuasive systems can support this.

When considering the third postulate, which deals with direct and indirect (or a combination of both) persuasion strategies, it may be difficult to determine which strategy to use. For example, direct persuasion might be more enduring than indirect persuasion strategies. However, an indirect strategy could be better for individuals who are in a hurry or in the event of information overflow via the persuasive system. In terms of SDT, it could be thought that the users should “know” and perform the right actions to achieve their goals, and persuasive systems can support this.

As stated in the fourth postulate of the PSD model, persuasion is often incremental, and therefore persuasive systems can enable users to proceed toward the target behavior through a series of incremental steps [13]. By using a persuasive system, the users should be encouraged to take small steps at the beginning and then take larger steps toward the target behavior over the course of the use process. TTM is suitable for incremental persuasion, as it is inherently divided into different stages, with the first stage focused on preparing users to achieve their personal behavior change goals.

The fifth postulate stresses system transparency, whereas the sixth postulate emphasizes unobtrusiveness [13]. If the system is biased with false information or disturbs the users, the outcomes in terms of behavioral change may be less than desirable. The mHealth app was backed up by a trusted national institute (the Finnish Institute of Occupational Health) and was designed to be unobtrusive. However, the experiences of some users may have varied due to bugs, specifically push notifications triggering at less-than-ideal times.

The seventh postulate of the PSD model, regarding the usefulness and ease of use of a persuasive system, indicates that useless systems or ones that are difficult to use are not that persuasive [13]. If the software quality of a system is poor or lacking, there is a high possibility that the system will not be used for a long time or continuously by the users. However, the situation is not specific to persuasive systems; rather, it applies to all information systems. Therefore, poor usability or bugs might reduce the use and thus the overall persuasiveness of any system.

The System

Overview

As the app was developed with the help of the PSD model, we analyzed and then selected the persuasive features to be used together with the research consortium. A workshop was held within the consortium at the beginning of the whole project, where principal investigators and researchers eventually chose the initial features based on reflections regarding the target group, previous experiences from similar research settings, and the trial context.

Furthermore, persuasive features were discussed with representatives of the target group in a series of focus group meetings and workshops [33]. During consortium meetings, the final set of features was eventually formed through discussions on what could support the target group, with background theories taken into consideration.

PSD Features

The persuasive features included in the app were based on the following PSD principles [13]: self-monitoring, rehearsal, praise, reminders, suggestion, liking, trustworthiness, and social comparison (Table 1).

Table 1. Principles that were implemented in the system.

<table>
<thead>
<tr>
<th>System support category</th>
<th>Principle</th>
<th>Example from the app</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary task</td>
<td>Self-monitoring</td>
<td>Step counter</td>
</tr>
<tr>
<td>Primary task</td>
<td>Rehearsal</td>
<td>Cyclic nutrition rehearsal tool</td>
</tr>
<tr>
<td>Dialogue</td>
<td>Praise</td>
<td>Positive feedback</td>
</tr>
<tr>
<td>Dialogue</td>
<td>Reminders</td>
<td>Push notifications</td>
</tr>
<tr>
<td>Dialogue</td>
<td>Suggestion</td>
<td>Pop-up giving a suggestion for behavior</td>
</tr>
<tr>
<td>Dialogue</td>
<td>Liking</td>
<td>Visually attractive pictures</td>
</tr>
<tr>
<td>Credibility</td>
<td>Trustworthiness</td>
<td>Evidence-based information</td>
</tr>
<tr>
<td>Social</td>
<td>Social comparison</td>
<td>Module proposition based on all users’ answers</td>
</tr>
</tbody>
</table>
Health Domains

Another paper by Laitinen et al \[34\] describes the study protocol of the randomized controlled trial, including the hypothesis behind the following health problem domains in the app: (1) exercising (physical activity), (2) stress management, (3) time management (efficient working hours), (4) recovery from work, (5) sleep, (6) healthy nutrition (dietary behavior), and (7) sedentary behavior (excessive sitting). In addition, the work by Tiitinen et al \[35\] was important to our choice of health domains. Figure 2 presents an edited screenshot of the app (textual content originally in Finnish but translated for this paper).

The results regarding the primary and secondary outcomes for the randomized controlled trial will be published in a separate paper in the future.

Figure 2. Health problem domains in the app.

App

We developed the system for the Android smartphone platform as a native app, which means that it was implemented using a compiled language (Java in this case) instead of web technologies \[36\]. Given the number of different Android smartphone devices available from various manufacturers and the differences within the Android operating system versions, the development process might have been less resource consuming using web technologies. Nevertheless, the native app approach could support use even without a network connection, thus enabling its use in remote locations, which we felt was an advantage over web-based apps.

Some software features implemented were more complicated in terms of programming, such as the step counter, whereas some were relatively simple in design. We used library packs provided by Android for the step counter; however, it took a relatively long time to test the functionality while adjusting the step counter. Thus, for resource reasons, we do not recommend adding complex tools to research purpose apps, as they are easily available elsewhere, and programming one from scratch (even with library packs) may require considerable time and resources.

The functionalities were designed to be simple and easy to implement. For example, we added a pop-up for certain intervals that provided relevant tips (per health module) to users. Similar to other tools, we strived for a simple yet efficient design, keeping in mind that we were designing a research app, not a finished and polished commercial product.

Another example of a simple design was the sit-stand reminder tool with an alarm. Although it would be possible to use the native alarm clock of one’s phone, it required relatively few resources (programming hours) to add one in the app. Thus, the users could use the tool easily within the app, as they were already committed to the trial. It would have served no purpose to ask users to find and install simple tools on their phones.

Lessons Learned

The lower limit of the Android version for using the app was 4.4, with no upper limit. The latest Android version available at the time of the intervention was 8.1. In hindsight, it would have been better to start development with the latest version, as the development took more than a year, and thus versions 4.4 up to 6 were already becoming outdated when the trial began. Too much variety in the Android versions increased opportunities for bugs because different Android versions, for example, use different libraries, and all differences between the versions had to be taken into consideration before the release of the app.

We should also have anticipated that many of our target users would have the latest smartphones and thus the latest Android
versions. In Finland, entrepreneurs can deduct the cost of work phones from their taxes. However, we also wanted to be fair and include earlier Android versions because not all microentrepreneurs can afford new phones. Indeed, self-employed entrepreneurs may live from “paycheck to paycheck,” only able to pay themselves salary depending on their sales or the number of customers in a given month.

Methods

Recruitment

Our research consortium recruited microentrepreneurs for the intervention via various means, such as email, and >1200 eligible participants were enrolled to participate in the randomized controlled trial. The recruitment process has been described in detail in another study [37].

The Trial

The enrolled participants were randomized into 2 groups: one for the actual intervention (613 participants) and another for control (612 participants). The control group was granted access to the same app with the same features at a later date than the intervention group. All participants were instructed to freely choose any of the health domains in their preferred order, and the app offered them the information and tools to reach their individual goals. In addition, they could use reflective questions in the app to determine their current situation regarding the health domains. We also informed them that they could freely perform any tasks within their preferred categories or just do them partially and they could always return to the tasks later. The participants were not compensated for participating in the intervention.

Although the intervention period was 8 weeks, the users could continue their use freely even after that period ended. The trial was conducted in Finland, and the participants eligible for the intervention had to live in Finland during the intervention and understand Finnish. Using an Android-based smartphone was essential and compulsory for participating in the intervention as well as being an actual microentrepreneur with fewer than 10 employees and financial revenue of €2 million (US $2.1 million).

All the interviewees in this study comprised of the intervention group. The full protocol of the trial is reported in another paper [34].

Data Collection

The interviews were based on semistructured questions in Finnish with 2 different emphases: the system (health behavior change, user, and use experiences; question set 1; Multimedia Appendix 1) and recovery from work (microentrepreneurs’ health and ways of living and app use for recovery; question set 2; Multimedia Appendix 2). The responses to the questions and discussions during the interviews were used to form the data sets: data set 1 (the system) and data set 2 (recovery from work). Although the angles varied in the interviews, the topics overlapped, and thus both data sets included discussions on similar matters.

We decided to use both data sets for this study as they complement each other, which leads to a more complete picture of the phenomenon under study. The questions for both interviews were piloted with the representative users before the trial.

Participants from the intervention group who had given their consent to be contacted were randomized into 2 lists for the interviews, and each of the 2 research teams responsible for the interviews received one list. The participants were contacted an equal number of times to obtain their final consent for the interviews and to schedule them. The first data set consisted of the interviews of 29 participants, whereas the second data set consisted of the interviews of 30 participants. Thus, a total of 59 interviews were conducted in this study.

The interviews were mainly conducted using Skype, a voice-over IP software program. Telephone calls were offered as an alternative in the event that participants could not use Skype for some reason. The recorded interviews were transcribed manually by third-party professionals. The contents of the interviews were not altered during the transcription process.

Ethics Approval, Informed Consent, and Participation

All the participants provided informed consent and were informed of their ability to opt out. The participants were interviewed as part of the randomized controlled trial study, which was approved by the Ethics Committee of the Finnish Institute of Occupational Health in November 2017 (#5/2017).

Research Methodology

We decided to use thematic analysis as the research method. According to Braun and Clarke [38], it is “a method for identifying, analyzing, and reporting patterns (themes) within data. It minimally organizes and describes your data set in (rich) detail.” There were 6 phases in the thematic analysis [38,39] (Figure 3).

Our data corpus consisted of the interviews, with each interview being a data item. Data extracts were individually coded chunks of data that were identified within and extracted from data items [38].

To accurately answer our 2 research questions, we performed 2 analyses. For the first analysis (persuasion event context analysis), we chose an inductive (data-driven) approach to identify the factors that hinder or reduce the use of the app. Thus, we did not attempt to fit the data based on preexisting frames or preconceptions. Direct implications for theory are not the priority in a data-driven qualitative data analysis and may not even be required. However, the theoretical implications cannot be fully ignored [38].

For the second analysis (PSD analysis), we used a deductive (theory-driven) approach, allowing us to compare the interviewees’ perspectives against the PSD framework model to see how we could mitigate the factors according to the interviewees’ views. The results of both analyses are further discussed in the Discussion section.
Data Analysis Process

**First Phase**
The transcribed interviews were carefully read 3 times to get familiar with the data [38,39]. A reflexivity journal [39] was initiated at this point in the form of memos and notes [38]. The journal was updated constantly throughout the analysis process. The first phase was similar in both the analyses.

**Second Phase—Persuasion Event Analysis**
Computer software can be helpful for the coding process [40]. Therefore, the transcribed interviews were imported into NVivo (QSR International), a qualitative analysis program. We generated initial codes from the data [38,39], which were divided into 2 deductive categories: (1) technical reasons reducing the use of the app and (2) other reasons reducing the use. Combining deductive and inductive approaches in thematic analysis is not uncommon and may be used when necessary [39,41,42].

**Second Phase—PSD Analysis**
The transcribed documents were imported into NVivo for coding. Initial codes were created and divided into four deductive theme categories according to PSD: (1) primary task support, (2) dialogue support, (3) system credibility support, and (4) social support. A theory, framework, or model can be used for creating deductive categories when performing a deductive analysis [38].

**Third Phase—Persuasion Event Analysis**
In this phase, we started to search for themes [38,39] from the codes in the initial coding categories. We identified 10 potential themes from the codes. These themes were moved into separate theme nodes (with work-in-progress names) in NVivo, deriving the following candidate versions of the themes: other apps, busy, content, Hawthorne, format, disappointment, usability, stress, life, and bugs.

**Third Phase—PSD Analysis**
Next, we formed subthemes for each main theme (PSD category) from the PSD model using category-related PSD principles [13] as subthemes. The codes were moved into equivalent or suitable subtheme nodes that best matched the codes. In this phase, we found that some codes seemed to overlap with the PSD principles; therefore, these codes were placed into ≥2 subtheme nodes at the same time for later decision-making in the next phase.

**Fourth Phase—Persuasion Event Analysis**
In this phase, we reviewed and refined the themes carefully [38,39]. First, we identified 2 general main themes (parent nodes in NVivo) from the candidate themes: “technology-related” reasons and “user-related” reasons that reduced and hindered the use of the app. The technology-related themes were linked to the app itself, whereas user-related themes were naturally linked to the users themselves.

However, we noticed that some subthemes found in the inductive analysis did not fit either theme. Thus, we added a third main theme: “use-related” reasons. As part of the review process, unnecessary and overlapping codes were deleted [38,39]. To tie the inductive analysis into the theoretical commitment [38], each refined subtheme was placed under the matching main theme (by switching to a deductive approach) with the help of the PSD definition of the persuasion event context regarding use, users, and technology. However, the subthemes remained relatively broad. Hence, we divided each subtheme into smaller nodes to highlight these issues in more detail.

**Fourth Phase—PSD Analysis**
We then determined the final subtheme placement for overlapping codes and deleted duplicates from other subthemes. We also noticed that some subthemes were either empty or the data we had for them were not rich enough for certain results, which may often happen [38]. In such cases, themes were removed from the analysis, but not before we returned to the raw data [39].

**Fifth Phase—Persuasion Event Analysis**
In this phase, the final themes were defined. We adapted the main themes under matching persuasion event context themes: use context, user context, and technology context [13]. Subthemes, including smaller nodes within the subthemes, were also given their final names [38,39] while trying to avoid refining the themes for too long. A reflexivity journal (memos
and notes) was maintained during the entire process, which involved writing about the analysis of the themes.

**Fifth Phase—PSD Analysis**

We went through the data and coding 2 times in this phase to ensure that we could develop credible results for the final analysis [39,40]. Similar to the persuasion event analysis, we completed the final analysis of the themes with the help of the reflexivity journal.

**Sixth Phase**

The last phase consisted of producing final versions of both the analyses and reports [38,39] using direct quotes. The reflexivity journal provided support in writing the analyses and report, which are presented as results in this paper. The entire research process (Figure 3) was time-consuming, but every phase was needed to conduct a thematic analysis [38]. Although thematic analysis can yield interesting qualitative results, the analysis process can sometimes be long and complex.

**Results**

**Persuasion Event—Use Context**

**Overview**

The lack of time due to work was the most common subtheme, and 71% (42/59) of the interviewees indicated that they were spending a lot of time working. We also found it to be linked with other subthemes. For example, the interviewees expressed that if the use was complex, they did not want to spend time on the learning curve, thus abandoning the app in the worst-case scenario.

A similar example of the lack of time reflected in other subthemes was that many interviewees felt that they were too busy to read the instructions or module introductions properly. Therefore, they may have concluded that the app was malfunctioning. However, the backend system that logged data from user interactions with the system offered a different view (although there were bugs in the system, not all “malfunctions” were bugs according to the log data). We think that some interviewees assumed that the app was malfunctioning when it may have been that they did not have time to learn to use the app properly (or to read the instructions).

The idea that technology is not working could also stem from technostress. In fact, technostress was a major subtheme within the interviews, as 29% (17/59) of the interviewees expressed feelings of stress due to the use of technology.

Furthermore, content-related issues were very common, and 59% (35/59) of the interviewees expressed having some issues with the content. It should be noted that the content of the app was somewhat extensive, and thus it is not surprising that this would arise as one of the top reasons for reducing app use. The results for each subtheme regarding use context themes are presented in Table 2.

<table>
<thead>
<tr>
<th>Use context themes</th>
<th>Interviewees (data set 1; n=29), n (%)</th>
<th>Interviewees (data set 2; n=30), n (%)</th>
<th>All interviewees (N=59), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lack of time due to work</strong></td>
<td>24 (100)</td>
<td>18 (100)</td>
<td>42 (100)</td>
</tr>
<tr>
<td>Busy at work</td>
<td>24 (100)</td>
<td>18 (100)</td>
<td>42 (100)</td>
</tr>
<tr>
<td>Excessive working hours</td>
<td>3 (12)</td>
<td>5 (28)</td>
<td>8 (19)</td>
</tr>
<tr>
<td><strong>Content-related issues</strong></td>
<td>22 (100)</td>
<td>13 (100)</td>
<td>35 (100)</td>
</tr>
<tr>
<td>Information overflow</td>
<td>10 (45)</td>
<td>7 (54)</td>
<td>17 (49)</td>
</tr>
<tr>
<td>Need for advanced content</td>
<td>13 (59)</td>
<td>3 (23)</td>
<td>16 (46)</td>
</tr>
<tr>
<td>Contents not suitable</td>
<td>5 (23)</td>
<td>4 (31)</td>
<td>9 (26)</td>
</tr>
<tr>
<td>Too much textual content</td>
<td>3 (14)</td>
<td>3 (23)</td>
<td>6 (17)</td>
</tr>
<tr>
<td><strong>Technostress</strong></td>
<td>11 (100)</td>
<td>6 (100)</td>
<td>17 (100)</td>
</tr>
<tr>
<td>Invasive technology</td>
<td>5 (45)</td>
<td>4 (67)</td>
<td>9 (53)</td>
</tr>
<tr>
<td>Stressed from using the app</td>
<td>3 (27)</td>
<td>2 (33)</td>
<td>5 (29)</td>
</tr>
<tr>
<td>System too complex</td>
<td>3 (27)</td>
<td>0 (0)</td>
<td>3 (18)</td>
</tr>
</tbody>
</table>

**Lack of Time Due to Work**

In one way or another, 71% (42/59) of the interviewees expressed being busy due to work. Furthermore, 19% (8/42) of them even reported working 7-day weeks or 15-to-16-hour days. Excessive working habits can affect life in several ways, for example, leaving little or no time to spend with family. Therefore, it is unsurprising that using the app might not be the first thing that the interviewees did when they had little time for themselves. In some cases, they even felt guilty for being too busy to use the app:

*I feel guilty, because using the app would not have taken that much time, no need to inspect everything for hours, so it would have fit [into daily routines] and I could have done something every day. I was so busy then, but now when I’m not that busy anymore, I have actually used the app more. [Interviewee #3]*

Many interviewees expressed that work came first, as they felt that they were responsible not only for themselves but also for their families and for their employees and their employees’ families. They seemed to be interested in changing their poor health choices to healthier ones, but they often neglected
themselves and their own health because they were busy. For 71% (42/59) of the interviewees, being busy due to work clearly reduced the use of the app, as they prioritized working over the app or even over everything else:

*It was four p.m., after which I used to work for four more hours, and I didn’t have time for anything else. My social life was suffering, and I spent the weekends at work. When there [from the app] came those reminders, I was at work. I just always ignored the reminders, because I never had any time to use it [the app].* [Interviewee #55]

The interviewees reported that during the intervention period they had a lot of work, it was their best seasonal time for working, or that they had to prepare for the coming season. The interviewees seemed to have more work available than they could complete within “normal” office hours. This resulted in tiredness due to long workdays, which left no time for anything else:

*At the beginning of spring, I was having this contract job that had been going on already for few months. It required me to drive tens of kilometers every morning, after which I did a long day and drove back. I was very tired, and I thought that this must stop, or I’ll stop being an entrepreneur. I was so tired, and I had no time for anything else [than work].* [Interviewee #19]

It was not just the actual work that caused the interviewees to be busy. They also expended considerable effort in obtaining contracts, jobs, or orders as well as in other work-related tasks, such as financial management and replying to customers. As microentrepreneurs, the interviewees also managed their own companies and possibly even had employees to manage, which led to more working hours. They also reported continuing working at home after they had left the workplace for the day:

*I don’t have time to get everything done during the day, so I work at evenings and nights too. I might go to bed at the same time as my kid, but then I wake up during the small hours to work, or I work at midnight.* [Interviewee #32]

**Content-Related Issues**

Overall, 27% (16/59) of the interviewees believed that the information provided by the app was general, with little new information to offer. Furthermore, 10% (6/59) of the interviewees thought that there was too much textual information to read:

*If that wall of text is even necessary, and this felt somewhat like lectures in the app, I really don’t know which kind of people even need that.* [Interviewee #5]

The information content was clearly problematic for some of the interviewees. They complained that they were unsure if the app was meant for them. For example, there were some tasks they felt that they could not complete, such as talking to colleagues (when one was working alone). Given the extensive amount of content, it was inevitable that some aspects of the content might be problematic for some users—for example, if the user did not perform office work with excessive sitting but was advised to stand up periodically. Better personalization could have solved this issue.

However, this was not a problem for all of the respondents, as they reported that they went through only the parts of the content that they needed. Even so, the wide scope of the app content presented a problem for 29% (17/59) of the interviewees in the form of information overflow, as they were unsure which health problem module or tasks and tools to pick:

*It takes a lot of time [to use] and last winter [time of the use] I was often very tired, so I found it hard to concentrate on these things here, because there are so much content and different modules.* [Interviewee #15]

**Technostress**

We noticed that when the users were busy because of work, they were also stressed because of work. Smartphones were seen as one of the tools for working, which caused stress. During busy periods, the interviewees thought that their phones were ringing “all the time,” and they also felt that they had to answer the phone when a customer was calling. Adding technostress to the equation of excessive working and being tired seemed to increase their perceived stress. Simply having to use a smartphone was named as a stressor in addition to receiving push notifications (reminders) from the phone.

Technostress, the inability to adapt to rapidly deployed new technologies, may have physical consequences for users, such as headaches, restlessness, or fatigue [43], or increase stress hormone production [44]. Overall, 8% (5/59) of the interviewees reported that using the app caused technostress for them:

*It was stressful and that I was supposed to be a slave to the phone even more, when I was thinking that I don’t want to check my phone all the time. It was the third day when I uninstalled the app.* [Interviewee #19]

**Persuasion Event—User Context**

**Overview**

In the user context, concurrent use of wearables or another app was the most common theme, as reported by 37% (22/59) of the interviewees (Table 3). It should be noted that while some users decided to use the alternative that best suited their needs, others chose to continue this concurrent use until the end of the intervention period. Other themes stemming from the users (which most people could relate to) were disadvantageous life situations, including health conditions, unfulfilled expectations (eg, vague “need” for something that is lacking), and different coaching preferences (eg, personal trainer instead of app).
Table 3. User context themes found in the analysis.

<table>
<thead>
<tr>
<th>User context themes</th>
<th>Interviewees (data set 1: n=29), n (%)</th>
<th>Interviewees (data set 2: n=30), n (%)</th>
<th>All interviewees (N=59), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concurrent use of wearables or another app</td>
<td>10 (100)</td>
<td>7 (100)</td>
<td>17 (100)</td>
</tr>
<tr>
<td>Using another application at the same time</td>
<td>6 (60)</td>
<td>5 (71)</td>
<td>11 (65)</td>
</tr>
<tr>
<td>Using wearables at the same time</td>
<td>7 (70)</td>
<td>3 (43)</td>
<td>10 (59)</td>
</tr>
<tr>
<td>Disadvantageous life situation</td>
<td>1 (100)</td>
<td>12 (100)</td>
<td>13 (100)</td>
</tr>
<tr>
<td>Medical condition</td>
<td>0 (0)</td>
<td>6 (50)</td>
<td>6 (43)</td>
</tr>
<tr>
<td>Mental health issues</td>
<td>0 (0)</td>
<td>4 (33)</td>
<td>4 (31)</td>
</tr>
<tr>
<td>Changes in everyday life</td>
<td>1 (100)</td>
<td>1 (8)</td>
<td>2 (15)</td>
</tr>
<tr>
<td>Family-related issues</td>
<td>0 (0)</td>
<td>3 (25)</td>
<td>3 (25)</td>
</tr>
<tr>
<td>Expectations unfulfilled</td>
<td>6 (100)</td>
<td>5 (100)</td>
<td>11 (100)</td>
</tr>
<tr>
<td>Expectations did not match with reality</td>
<td>2 (33)</td>
<td>3 (60)</td>
<td>5 (45)</td>
</tr>
<tr>
<td>Lacking features</td>
<td>2 (33)</td>
<td>2 (40)</td>
<td>4 (36)</td>
</tr>
<tr>
<td>Lacking something</td>
<td>2 (33)</td>
<td>0 (0)</td>
<td>2 (18)</td>
</tr>
<tr>
<td>Different coaching preference</td>
<td>4 (100)</td>
<td>6 (100)</td>
<td>10 (100)</td>
</tr>
<tr>
<td>Face-to-face preferred</td>
<td>1 (25)</td>
<td>3 (50)</td>
<td>4 (40)</td>
</tr>
<tr>
<td>Nondigital self-help preferred</td>
<td>3 (75)</td>
<td>2 (33)</td>
<td>5 (50)</td>
</tr>
<tr>
<td>Medical measurements preferred</td>
<td>1 (25)</td>
<td>0 (0)</td>
<td>1 (10)</td>
</tr>
<tr>
<td>Peer support groups preferred</td>
<td>0 (0)</td>
<td>1 (17)</td>
<td>1 (10)</td>
</tr>
</tbody>
</table>

**Concurrent Use of Wearables or Another App**

Overall, 29% (17/59) of the interviewees used wearables or another app during the intervention period, often concurrently with the persuasive mHealth app used in the trial. Wearables (eg, smartwatches) were popular as self-monitoring tools, for example, for measuring the user’s steps or heartbeat. One reason for using wearables was that people might leave their phones on their desks while walking to the printer. In contrast, wearable devices could be carried easily without any extra effort:

> I don’t carry my phone with me all the time, and the app assumed that everyone would carry her or his phone everywhere. I have a smartwatch, which I use for measuring my steps and pretty much for everything else, too. [Interviewee #5]

Users who were not interested in reading or who disliked the coaching approach of the app seemed to find wearables or simple sport apps better suited for their needs. They seemed to be mainly interested in measuring different health-related aspects, such as heartbeat or sleep, and were less interested in being coached. However, they could use the trial app concurrently with the wearables or another app:

> I have recently installed another app, which I use for following what I eat, but otherwise I don’t have anything else related to health in my phone. Oh, but wait, I do have an activity band too, from which I get data into my phone, and then there is the app of yours. I have noticed that these are helpful for checking things out. [Interviewee #16]

**Disadvantageous Life Situation**

People tend to experience different situations in their lives, which could hinder or decrease the use of any app. The microentrepreneurs interviewed were no different, as they were troubled by loud neighbors or experienced insomnia, insecurities regarding their business, health issues (or their relatives had health issues), etc. For some people, combining entrepreneurship and family life can be difficult, as both might require a considerable amount of time. As the app was dealing with health problem domains, it was unsurprising that 17% (10/59) of the interviewees reported experiencing either mental or physical health conditions, which reduced their use of the app:

> It is probably because of my condition, as I can’t concentrate on anything in the kitchen. That section, “plan your meals” in the app, well that planning thing, as well as putting it into practice, is difficult for me because of my condition. [Interviewee #50]

**Expectations Unfulfilled**

Overall, 19% (11/59) of the interviewees had great expectations for the app but were disappointed in practice. In other words, their expectations did not necessarily match the reality of using the app. If users have predefined needs, and if they think that they cannot fulfill those needs with the help of the app used, they will surely be disappointed. This will inevitably reduce their use of any app.

The interviewees found it difficult to point out exactly what it was that they were missing, but they mentioned issues such as:
networking (with other entrepreneurs), peer support, and various automated measurement functions:

I didn’t find what I was looking for, although clarifying what I wanted is difficult, but I thought that it could have automatically offered what I needed. I cannot really put it into words, just a thought in my head, but it should have measured me automatically during the workday, like how much I am sitting or how stressed I am, or other stuff like that. Pretty tough demands and so on. [[Interviewee #20]]

Different Coaching Preference

Apparently, not all users knew what they were enrolling in, although the intervention was advertised as the use of an evidence-based coaching app for behavioral changes. For example, 3% (2/59) of the interviewees complained about the chosen coaching approach, stating that they would have preferred to see a health care professional face-to-face. Two other interviewees resorted to hiring someone to help them (eg, a personal trainer), which led them to abandoning the app:

I have glanced at the contents, but it didn’t inspire me that much, because I had a chance for this hired personal guidance face-to-face. In my opinion, an app can’t compete with humans yet, and I managed to get expert guidance otherwise. [Interviewee #34]

Persuasion Event—Technology Context

Overview

Of the 59 interviewees, 30 (51%) complained of technical issues. It should be noted that being busy at work (no time to learn to use the app or read instructions) could have affected this perception. We do not disagree with their views (as views tend to be subjective experiences), but we do conclude from log data that not everything reported as a technical error was one. However, if the interviewees felt that there were technical issues, then it does not matter whether they were real. Better usability (considering both the use and user contexts) could have solved this issue, at least to some level.

Usability issues were common factors hindering the use, and 37% (22/59) of the interviewees reported such issues. Usability was also partially tied to content-related issues. As a background theory, SDT affected the usability and content of the app. When designing the app, we assumed that people would make the “right” choices most of the time (from the given options).

Some users perceived it difficult to choose a health problem domain from the options while also having to choose the proper stage of change when wanting to perform tasks from the modules. In addition, 17% (10/59) of the interviewees did not like the platform that was used (native Android app; Table 4).

Table 4. Technology context themes found in the analysis.

<table>
<thead>
<tr>
<th>Technology context themes</th>
<th>Interviewees (data set 1; n=29), n (%)</th>
<th>Interviewees (data set 2; n=30), n (%)</th>
<th>All interviewees (N=59), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technical issues</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major bugs</td>
<td>5 (25)</td>
<td>3 (30)</td>
<td>8 (27)</td>
</tr>
<tr>
<td>Minor bugs</td>
<td>16 (80)</td>
<td>7 (70)</td>
<td>23 (77)</td>
</tr>
<tr>
<td>Usability issues</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difficult learning curve</td>
<td>10 (71)</td>
<td>5 (62)</td>
<td>15 (68)</td>
</tr>
<tr>
<td>Complexity issues</td>
<td>5 (36)</td>
<td>4 (50)</td>
<td>9 (41)</td>
</tr>
<tr>
<td>Memorability</td>
<td>1 (7)</td>
<td>1 (12)</td>
<td>2 (9)</td>
</tr>
<tr>
<td>Disfavored platform</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Different format preferred</td>
<td>2 (67)</td>
<td>6 (86)</td>
<td>8 (80)</td>
</tr>
<tr>
<td>Different operating system preferred</td>
<td>1 (33)</td>
<td>1 (14)</td>
<td>2 (20)</td>
</tr>
</tbody>
</table>

Technical Issues

In the interviews, many users complained about numerous but mostly minor issues with the app (eg, incorrect font size and screen scale). One crucial bug related to persuasiveness was that the weekly push notification reminders did not work for all users. Overall, 25% (15/59) of the interviewees reported that they did not receive weekly reminders, or that when interacting with the weekly reminder and trying to answer the questionnaire, they could not submit the answer:

That recovery statistic reminder or something like that, there was a bug, since even after two weeks when trying to input the answer, the program replied instantly that I had updated the answer already in that week, and I should try again later. [Interviewee #54]

On one hand, if they did not receive the weekly reminder, participants reported that they forgot to use the app or the feature with the malfunctioning reminder. On the other hand, when they received a malfunctioning weekly reminder, it also decreased the persuasiveness and use of the app owing to user frustration or disappointment because of the bug.

For 8% (5/59) of the interviewees, the push notifications occasionally malfunctioned and looped the weekly reminder or task reminders unnecessarily. In a rare case, a large question set (supposed to be triggered at the end of the intervention period via push notification) was looped, which eventually led the interviewee to abandon the app:
That set of questions was long and so I tried to proceed from it, answered the questions and accepted them, ok. This kind of app should be intuitive, so nothing is left hanging. I think I made it to the end and continued from there, where there were these tasks and picked few of them to start with. The next time I used the app, it wanted me to do the question set again, and I did not want to do that. It offered it to me at least three or four times. [Interviewee #26]

In addition, 7% (4/59) of the interviewees reported that the app either froze or crashed on their smartphones occasionally, but otherwise the bugs mostly hindered rather than prevented use. Nevertheless, any bug, whether minor or major, might reduce the persuasiveness and use of any app. On the one hand, users might wonder if it is worth continuing to use an app that does not seem to work properly—there are many alternatives in the commercial market. On the other hand, bugs might be something that more experienced users have become used to, at least to some extent, as one interviewee expressed:

Oh well, it must be because usually all of these [health apps] don’t necessary work, so I’ve gotten used to it that these just happen to have these [bugs]. [Interviewee #3]

Usability Issues
All 22 (37%) interviewees in this theme were either unsure about how they should have used the app or felt that the app was too complex. They complained that the learning curve was too high and that there were no clear instructions on how to use the app (or that they could not find the instructions):

When going through the app, I thought that there would be instructions on how to use it, how it works, what is the idea behind it, but I didn’t find anything like that. A month later, I think, I found instructions from somewhere, which explained a little. [Interviewee #7]

It should be mentioned that when logging into the app for the first time, there were instructions on how to use the app and the concept behind it, but some users skipped the introduction. The same introductory text was also available under the main menu. In addition, the instructions on how tasks work were available each time the user chose a task.

Nevertheless, when users felt that instructions were lacking, it reduced their use because they were unsure of how to use the app. At a general level, some people may be irritated by excessive explanations and instructions, whereas others may quit using apps because of a lack of clear and plentiful instructions. The interviews also showed that another usability flaw from the users’ viewpoint was the lack of an option to check which tasks had already been performed:

I want to see my progress, so in that sense, for example in tasks there is no list of what I have already done, or anything like that, where I could check on how the task went. [Interviewee #3]

Disfavored Platform
Overall, 7% (4/59) of the interviewees did not like using the app on the smartphone and would have preferred alternatives. One of them reported that because of a medical condition, a keyboard and a mouse would have been a better option, as handling a touchscreen on a smartphone was painful. Another would have preferred a radio broadcast (podcast) for guidance rather than a smartphone app. Activity bracelets and smart watches have also been mentioned as a preferred platform because they have better sensors and offer automatic measurement. Furthermore, 3% (2/59) of the interviewees complained that because only an Android version was available, they had to use it with their secondary phones, as they mainly used iPhones. This evidently decreased their use of the app:

I use iPhone, but I have one Android phone, into which I installed this app, because there was not an iPhone version available. I do not normally use an Android phone. Because of that, I haven’t used the app very much. [Interviewee 53]

It should be mentioned that dozens of iPhone users enrolled in the intervention, although it was clearly advertised that the app was available only for Android smartphones. The enrollment web form included a specific question about whether the users had an Android phone. If a potential participant answered that they did not have or use an Android smartphone, the enrollment did not continue. Apparently, these people either answered incorrectly to continue or did not read the question properly.

Moreover, our helpdesk was approached several times via email by iPhone users complaining about the lack of an iOS version. It is therefore possible that several participants switched from Android phones to iPhones between the enrollment phase and the start of the trial.

PSD Analysis—Persuasive Categories and Principles
Overview
Unsurprisingly, primary task support was the top PSD category in the analysis. Primary task principles support primary tasks, as indicated by the name. Something that was a bit surprising was that system credibility support emerged from the analysis, as it has been given less attention by both users and designers in the past. However, only 1 principle came up, and only with 3% (2/59) of the interviewees, so this was not a strong issue.

Dialogue support had 2 principles. Many of its features can be seen as supporting not only dialogue but also primary tasks. For example, reminder reminds the user to use a self-monitoring tool. Social support is another category that users may like in general.

In the analysis, only 2 social support category principles emerged from the interviews. Social features were present in the app, and they were discussed in the interviews; therefore, this result is likely related to the research question (how to mitigate hinderances) rather than a lack of interest.
Primary Task Support

Overview

Personalization was the top principle in this category, which was mentioned by 27% (16/59) of the interviewees. Tunneling was discussed by 19% (11/59) interviewees and self-monitoring by 17% (10/59). Tailoring was brought up in 8% (5/59) of the interviews and reduction only in 3% (2/59) (Table 5).

Table 5. Primary task support features found in the analysis.

<table>
<thead>
<tr>
<th>Primary task support</th>
<th>Interviewees (data set 1; n=29), n (%)</th>
<th>Interviewees (data set 2; n=30), n (%)</th>
<th>All interviewees (N=59), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personalization</td>
<td>16 (100)</td>
<td>1 (6)</td>
<td>16 (100)</td>
</tr>
<tr>
<td>Tunneling</td>
<td>7 (64)</td>
<td>4 (36)</td>
<td>11 (100)</td>
</tr>
<tr>
<td>Self-monitoring</td>
<td>7 (70)</td>
<td>3 (10)</td>
<td>10 (100)</td>
</tr>
<tr>
<td>Tailoring</td>
<td>5 (100)</td>
<td>0 (0)</td>
<td>5 (100)</td>
</tr>
<tr>
<td>Reduction</td>
<td>2 (100)</td>
<td>0 (0)</td>
<td>2 (100)</td>
</tr>
<tr>
<td>Simulation</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Rehearsal</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

Personalization

Personalization can be defined as providing personalized content or services [13]. Different ideas for personalization or even customization of reminders, menus, and content in the app emerged in the interviews. However, perhaps the most important finding regarding this theme was that all 16 (27%) interviewees felt that personalization would have improved their motivation and engagement with the app.

Given that personalization in the app only involved personalized suggestions, such as which health module to select in the beginning, it was unsurprising that the interviewees felt the app was lacking in this respect. Even “light” personalization without the participants being able to customize things would have been welcome:

Yes, so it could have taken into consideration what or who I am, my age, and the work I do and so forth. So, it would have been better if these would have been considered [in the app]. [Interviewee #12]

On the basis of the context analysis, many interviewees felt that the content was too general and at times felt that it was not meant for them. Personalizing the content, for example, according to the type of work done or the work environment, could reduce this issue and improve engagement. Thus, users who do not sit in front of a desk during workdays would not be encouraged to stand up regularly by the app. This type of “lighter” personalization could easily be accomplished with a few quick preuse questions (eg, “Do you work in an office environment?” or “Does your work require a lot of standing or moving?”).

The interviewees noted that many companies knew a lot about their users. For example, Google collects various data about users and their app use. The data that an app collects could then be used to personalize the app based on use patterns (eg, number of steps taken in certain periods).

Two interviewees went even further regarding their expectations of the app, suggesting self-learning algorithms:

It could be even more precise, yes it could, and I would say that artificial intelligence could be utilized, so it would match even more precisely into your own profile. [Interviewee #11]

Personalization could also potentially counter the need for concurrent use of other apps or even wearables, especially if the wearables could be synced to support the app and the collected data could be used. Indeed, personalization, by providing better correspondence between the app and the needs of users, could help to remove the motivation to use complementary apps. Personalization has been shown to be effective in supporting behavior change, but it is not used to its full potential in current mHealth apps [45].

Tunneling

Tunneling means that the system should guide users toward the target behavior [13]. We did not implement tunneling (predesigned use paths within the app in this case) because we interpreted the autonomy aspect of SDT strictly. We felt that it would be best to allow users as much freedom as possible in navigation, presuming that the users would then choose the “right” actions in the app.

However, with tunneling based on personalization, the predesigned use paths could have been based on users’ own choices, thus not contradicting SDT in that sense. The same interviewee who brought up artificial intelligence regarding personalization also spoke about tunneling based on personalization:

Well, so these [use] paths, I think that they good in the sense that depending on your situation you can take a certain path [of use]. Be it exercising, or mindfulness, or [healthy] eating, or what. [Interviewee #11]

Another interviewee articulated tunneling based on personalization in a more thorough manner:

Yes, a clear path which you follow so there won’t be too many options, because if you are at a crossroad and you have many paths to follow, you have to choose one, and then it may be difficult because you don’t remember which path you took. [With] one path, you can follow the tunnel to the end and only then take another, which would be so much clearer for me.
When you go home and start using the app, you are like what’s the deal, but those straightforward paths take less time, when you don’t have to search [what to do next]. [Interviewee #15]

It became apparent from the interviews that tunneling could also “hit two birds with the same stone.” This is because many interviewees were having difficulties with both lack of time (to use the app) and information overflow (due to the broad content). Personalized tunnels would save time, as users could just start using the app even if they had only a few minutes. This is because they would not have to start by “learning” or deciding what to do next; rather, they could just go along the program until they know what to do.

Furthermore, if the tunnels or use paths are based on users’ personal preferences, users will not be overwhelmed by a massive amount of information. Instead, they will be offered only the correct path to navigate. Tunneling could also improve usability, which was problematic for many interviewees, by reducing the learning curve:

_It doesn’t mean that it would necessarily have to guide you step by step, but it could repeat [for the user] the idea and what it holds, how it works, or how it should be used so you could understand. It’s the same if you have never driven a car before and you are put behind the feel with no idea or anyone saying what you must do, then it may be that you don’t succeed at the first time trying to drive._ [Interviewee #7]

### Tailoring

Tailoring is related to personalization, but it focuses on user groups instead of individuals [13]. Similar to personalization, tailoring could help to address the issue of mismatching content with group levels (eg, office workers, self-employed microentrepreneurs). Moreover, tailoring could also improve users’ motivation to use the app, for example, through a social comparison function in the app (there were 2 features that showed comparisons of the results of the whole user base) that has different target groups:

_It’s nice to see what kind of stress levels we micro-entrepreneurs have at certain times, but since there are so many different types of micro-entrepreneurs it is difficult to compare the results...It would have been better if there would have been like the entrepreneurs of the same line of business to check._ [Interviewee #28]

### Reduction

Reduction, that is, reducing the complex behavior in the system into smaller tasks on the path to the target behavior [13], was requested by 3% (2/59) of the interviewees. Further reduction could save users some time, especially if they are extremely busy. However, as the theme only came up in 2 interviews and the app had already undergone considerable reduction at several levels (goal setting, 3 kinds of tasks from quick to long, and easy-to-use tools), further reduction would likely have only a minor effect on improving user engagement.

With other apps, it might be more beneficial for the “medicine” for behavior change to be provided in “doses.” This would be especially useful if the users are busy, tired, or stressed as by focusing on small steps, designers can avoid overwhelming users’ cognition.

### Self-Monitoring

Self-monitoring means that the system should provide a means for tracking one’s performance or status [13]. The app provided several types of self-monitoring, including self-reporting levels of stress or recovery and a variety of tools (eg, an alarm to remind the user to stand up). However, there was only one self-monitoring tool that took advantage of smartphone sensors for “automatic” measurement, a step counter. On the basis of the interviewees’ statements, improving the “automatic” monitoring functions of the app (via sensors or even syncing external wearables) could help to increase user engagement:

_I thought that this app would remind me about it [going to bed], and via the app I could also measure like an engineer what it actually is [amount of sleep], so it wouldn’t just be gut feeling [how much I sleep]. In a way, it would be a motivator, that kind of monitoring tool, which would help me to see the direction I’m going to and do I have some difficulties regarding sleeping or not._ [Interviewee #29]

Technostress can be mitigated by controlling the way technology is used and by distancing oneself from technology use when feeling stressed [46,47]. By enabling automated self-monitoring (via sensors) or syncing wearables to the app, designers could actively reduce technostress for users.

Overall, 29% (17/59) of the interviewees reported symptoms of technostress, and 5 (29%) of them stopped using the app. This is an important issue that would also affect similar apps. Thus, mitigating technostress could have an important impact on engagement. This is especially important for target groups that use technology as a means of working—they may not want to use mHealth apps to recover from work or manage stress if the app use reminds them of their work:

_It should have had, well something like activity bracelet or other automation. For me, it proved out to be too big of an issue to type things on my smartphone, because then I get the feeling that I must do that too much already, so I just want to get rid of that [typing on smartphone]._ [Interviewee #42]

### Dialogue Support

#### Overview

In 19% (11/59) of the interviews, the interviewees’ opinions about dialogue support focused mostly on reminders. Three interviewees (5%) also brought up liking (Table 6).
Reminders

Reminders from the system can remind users of their target behavior when using the system [13]. The interviewees felt that reminders should be meaningful and even customizable, which in this paper is linked to personalization. One interviewee (2%) thought that only getting a weekly reminder would be sufficient, whereas 2 (3%) interviewees proposed a weekly reminder in the form of a weekly review in addition to other reminders:

*I would like this to be more active, it should be more active for the users in some way. Weekly review would be very good, or weekly reminder on it, then it would work really well.* [Interviewee #2]

Two (3%) interviewees indicated that they would have been satisfied with fewer reminders, whereas 15% (9/59) of the interviewees wished for more than they had received:

*It [low use] is partially because I didn’t realize how good it is [the app], so maybe in the beginning there should have been [more] reminders. Naturally, some may be irritated by those, if some program reminds that now you have taken 10,000 steps, but this could have reminders more like think about this or have you checked that.* [Interviewee #8]

It should be mentioned that the push notifications in the app did not work perfectly for all users, which could have affected why 15% (9/59) of the interviewees requested more reminders. However, to the best of our knowledge, the bug was caused by mismatched libraries owing to the different Android versions. Therefore, it only applied to certain timed weekly reminders and not all reminders. Those that could have been triggered by choosing longer tasks seemed to work better according to the use log data.

Liking

Liking implies that the look and feel of the system should appeal to users [13]. A multitude of visually attractive pictures were used, and attention was given to how the text was set up and sectioned in the app. However, according to 5% (3/59) of the interviewees, the same principle was not applied to all the infographs in the app:

*When some graphs like in the app comes along, for me these are like something that I bypass very easily, since I just think that I don’t understand these kinds of crooked objects, or I don’t want to concentrate on them.* [Interviewee #15]

Improving the visual design of the infographs—making them easier and clearer to perceive and understand—would likely help to solve some issues regarding usability and content. In general, if users have difficulty in understanding or even noticing some aspects of the app, they will use the app, or at least those aspects of it, less.

One user even mentioned during the interview that they had used the tool with an infograph but had not paid much attention to it. A graphic designer worked on other parts of the app, but in hindsight, she should have also checked the infograph designs before implementation.

System Credibility Support

Overview

Only 3% (2/59) of the interviewees mentioned principle or principles related to system credibility support (Table 7), which was unsurprising. Features in this category are more difficult to implement as distinct technical features in apps. Some of them are even concepts that people have grown accustomed to and assume to be part of every app. For example, the expertise principle states that mobile apps should be updated regularly [13], which is something that is performed in the background.

Therefore, the lack of mention of these principles in this analysis does not mean that they are not good features; rather, they were already present to a sufficient degree. For example, some users mentioned that having the Finnish Institute of Occupational Health involved was important, which could be linked to expertise or even authority and trustworthiness depending on user’s perspective.
Table 7. System credibility support features identified in the analysis.

<table>
<thead>
<tr>
<th>System credibility support</th>
<th>Interviewees (data set 1; n=29), n (%)</th>
<th>Interviewees (data set 2; n=30), n (%)</th>
<th>All interviewees (N=59), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real-world feel</td>
<td>0 (0)</td>
<td>2 (100)</td>
<td>2 (100)</td>
</tr>
<tr>
<td>Trustworthiness</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Expertise</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Surface credibility</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Authority</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Third-party endorsements</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Verifiability</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

Real-World Feel

Real-world feel can make it possible to contact specific people through the system [13]. In total, 3% (2/59) of the interviewees brought up a real-world feel as a potential improvement to the app. One of them even stopped using the app due to receiving face-to-face guidance in real world. We agree that mHealth apps could use real-word feel, for example, through chats or meetings with health personnel, or as one of the interviewees expressed it:

It could be for example a nurse who you would meet regularly so you would follow [your progress] together with the nurse and you would be moving forward [towards personal goal]. So, a continuous care or well-being relationship would be formed. Yes, something along that line. [Interviewee #54]

For users struggling with difficulties related to guidance (different coaching preferences), real-world feel in the app could improve engagement. However, it is unclear how realistic it would be to implement real-world feel in meetings (either “live” or internet based) if the purpose is to develop cost-efficient health intervention apps. Such a feature might work better with more specialized or highly commercial (pay-per-use) guidance apps.

Social Support

Overview

Overall, 5% (3/59) of the interviewees mentioned social learning. Two discussed social comparison, which already existed as a feature in the app (Table 8).

Table 8. Social support features identified in the analysis.

<table>
<thead>
<tr>
<th>Social support</th>
<th>Interviewees (data set 1; n=29), n (%)</th>
<th>Interviewees (data set 2; n=30), n (%)</th>
<th>All interviewees (N=59), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social learning</td>
<td>3 (100)</td>
<td>0 (0)</td>
<td>3 (100)</td>
</tr>
<tr>
<td>Social comparison</td>
<td>2 (100)</td>
<td>0 (0)</td>
<td>2 (100)</td>
</tr>
<tr>
<td>Normative influence</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Social facilitation</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Cooperation</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Competition</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Recognition</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

Social Learning

Social learning suggests that users will be more motivated if they can observe others engaging in similar behaviors via the system [13]. Overall, 5% (3/59) of the interviewees indicated a desire for a networking or social learning feature in the app, if nothing else, at least peer chat support. The participants in our target group seemed to be social or at least interested in networking. This may be because networking may lead to business opportunities, and one can safely let off the steam caused by entrepreneurship with peers:

For micro-entrepreneurs, self-employed entrepreneurs, or those who employ few people, they may have little connections or networks. For these kinds of people, they could use this kind of [feature in the app]. I don’t know, perhaps chat, so they could share things safely among themselves. [Interview #24]

Sharing issues, such as worries or job-related strains, allows users to see that they are not alone, and in the case of the trial app, to see that others are working on recovering from work as well as trying to change their behavior in a healthier direction. Enabling social learning (or even networking) could resolve issues related to the “expectations unfulfilled” theme, and the interviewees stated that networking or peer support should be a part of the app. It is possible that this feature could reduce the obstacles for some users regarding use, but the interviews did not provide enough data to determine whether this is an important issue with this type of health app in general.
Social Comparison
Social comparison enables users to share and compare meaningful information with other users, which can increase their motivation to perform the target behavior [13]. The app had 2 social comparison features, but they targeted the entire user base. Overall, 3% (2/59) of the interviewees stated that this feature was not useful because of the lack of distinct groups:

Well, it didn’t influence me because there are probably some many different kinds of people, so at least I couldn’t see a distinct trend from it [social comparison feature]. [Interviewee #4]

With tailoring, it could be possible to divide users into different groups and only show comparison data from the group that equates to user. This could increase motivation for some, but it is unclear whether this would address any of the themes found to hinder their use in this study.

Discussion

Principal Findings
This paper presents some unanticipated findings, but in hindsight they are logical. They also showed that background theories should not be interpreted too strictly, or at least designers should find ways around them.

For example, when designing the app, we felt that we could not use tunneling, a PSD feature that “guides” users via a path toward the desired behavior. This is because SDT was used as the background theory, and there was concern that tunneling might interfere with the “free will” (autonomy) of the users. Thus, in seeking to avoid restricting the users, we managed to alienate the users who wanted “tour guidance” in using the app and their behavior change process.

Dual Role of Microentrepreneurs
The microentrepreneurs in this study appear to play dual roles. They were representatives of their own business, which affected the specific work-related strains and stressors they encountered and thus their recovery from work. At the same time, they were entrepreneurs, resulting in another range of strains and stressors, especially for those with employees. The dual roles of target groups represent a design challenge. From this viewpoint, will they use the apps? Do they have time to use apps at all?

Because of these dual roles, it was not surprising that two-thirds (42/59, 71%) of the interviewees reported being very busy, which seemed to be characteristic of their lives in general. When planning a trial for people with dual roles (or designing apps for them), it is important to consider that they may not be willing to spend a lot of their time. They may already be busy with other tasks and have no time for anything “extra.”

Need for Time-Saving Guidance
Given that many interviewees reported being busy, the need for better and personalized guidance was evident from the interviews. It is logical that busy people would like to avoid learning curves (with the help of tunneling) and only spend time on things that are explicitly useful to them (with the help of personalization).

Using tunneling, tailoring, reduction, and personalization may improve engagement as each principle can involve time-saving elements in the right context. With tunneling, the path is laid out for the user, especially when combined with personalization, and thus the learning curve should be less steep. Furthermore, users are not required to decide what to do in the app; instead, they can follow the guidance based on personal preferences.

In addition, if reduction is used correctly, users can digest small bits of information when they have time and do not become stressed due to lack of time. Moreover, reduction can save users from trying to absorb the whole thing at once, which may lead to information overflow and dropping out from the guidance program. Personalization could also enhance the user experience, as it would allow personalized content, which would certainly be more meaningful than general information for the user.

Technostress
Overall, 8% (5/59) of the interviewees reported experiencing technostress during the trial and quit using the app. Although this was not a common occurrence, they quit the trial due to technostress. In addition, technostress manifested when users had to learn how to use the app, although it did not require an insurmountable effort.

We learned from the analysis that interviewees who were already stressed due to work did not like using the same platforms or devices for recovery that they also used for work. Self-monitoring tools that are synced to wearables (or that use smartphone sensors for automatic measurement) could help users who want to take measurements while distancing themselves from active smartphone use outside of office time to decrease technostress.

Learning to use the app required at least some effort to read the instructions, which might have been too much for some users, particularly if they were already exhausted. Therefore, due to the lack of time and job-related strain, this may have triggered further technostress in some interviewees. Furthermore, technostress was likely increased by the reminders, as the participants could not customize them. They could only turn off the reminders for each task after receiving the first push notification.

PSD Postulates
The sixth PSD postulate states that a system should not be obtrusive; in this case, personalization could have decreased the obtrusiveness of the app. Obtrusiveness was caused by reminders triggering at the wrong time (when the interviewees could not react to them because they were working). Therefore, designers should enable customization of push notifications in systems—or at a minimum the ability to turn them off.

The fifth PSD postulate emphasizes transparency. Accordingly, designers should disclose what their apps are based on. Some interviewees stated that they knew that there was the Finnish Institute of Occupational Health behind the app, and so they felt it was trustworthy. Thus, there should not have been any confusion about the app being used for research and that it was not a commercial one. Nevertheless, it seems that we could have
done better in informing people enrolled in the trial, as some interviewees clearly did not realize what they had enrolled in or presumed they would be using apps similar to commercial ones. This could reduce use, so designer bias (the app being for research) should be clearly disclosed to users to avoid confusion.

Regarding whether to use direct or indirect persuasion (the third PSD postulate), it seems clear that in the case of microentrepreneurs, the indirect approach is better. The participants were constantly busy with their work; two-thirds (42/59, 71%) of the interviewees did not seem to have enough time to use the system, which also meant that the persuasion process might not have affected them continually or even incrementally as intended (first and fourth PSD postulates).

In addition, information overflow seemed to be an issue, as the interviewees reported that they had difficulties deciding on what to choose within the app. Thus, an indirect approach might be better if there are several possible ways to use an app or if several health problem domains are addressed in a single app. Moreover, we recommend using the tunneling principle in similar cases, as it could help the users with the learning curve and save precious time, thus enabling the system to be more open (“always on”) and the persuasion process to be incremental because users actually use the system. In addition, this could increase users’ commitment (second PSD postulate) by making it easier to use the app. Intuitively, it is easier to commit to something that can be used with a “plug-and-play” mindset rather than something that requires a steep learning curve.

The information overflow and the steep learning curve reported by the interviewees might have partially been a result of our strict interpretation of the autonomy aspect of SDT, which led to giving the users excessive freedom when navigating the app. If we had (better) used personalization, the choices for “tunnels” or use paths could have been the users’ own, in which case there should be no contradictions with SDT.

The seventh postulate of PSD encourages the design of useful and easy-to-use apps. Therefore, researchers and designers should be realistic about the features and the content of mHealth apps. For example, small start-ups or smaller research projects may not have adequate resources to implement everything. Carefully drawn lines defining what can and cannot be done with the given resources would result in more stable apps and fewer bugs for users (or developers) to worry about.

Lessons Learned

We acknowledge that the usability of the app could have been improved, as is evident from the analysis. In addition to usability issues, bugs in the app also reduced use. Although half (30/59, 51%) of the interviewees reported encountering bugs, most of them did not contact our helpdesk for technical support. Apparently, providing technical support via email alone is not sufficient for bug reporting [48]. It should be noted that some of the bugs reported by the interviewees could have been usability issues rather than technical difficulties.

The expectations of the users regarding the app seemed to be at least partially based on commercial health apps, and some were even mentioned during the interviews. Commercial apps differ from the app used in the intervention. The contents of the intervention app were evidence based, and the app was based on behavior change theories. At least in part, this could explain why some people felt that the tasks were different than those of commercial apps. Furthermore, based on the interviews, people have become accustomed to commercial apps having bugs, which are fixed eventually.

When enrolling in the intervention, not all participants may have had clear personal goals. Some may have joined simply out of curiosity, wanting to test the app. If it did not seem to suit their needs immediately or they felt it was too complex, they might have just abandoned it and moved on to the next one, and there are plenty available in the commercial market. Therefore, there is no need to try to engage 100% of users, as some people may just want to test it and may not be ready to engage.

We also recognize that it is not always an easy task to prioritize features in the design phase, and target users may end up behaving differently regarding app use than the designers originally predicted. Therefore, it is important to increase the knowledge about different user groups. However, it is not practical to try to meet every imaginable need of users, as there will always be some who will not be happy. Indeed, trying to fit everything into a single app may lead to poor design or imperfect implementation, thus benefitting no one.

Limitations

The limitations of this study include the differences between the data sets, as the 2 teams conducting the interviews used different sets of semistructured questions. The emphasis of the interviews was also different between the teams, although the themes of the actual questions overlapped in both data sets. The results can be generalized to similar groups to a certain extent, and a persuasive event analysis would be helpful for identifying those groups. However, it is also possible that different results could be obtained with similar groups.

The thematic analysis process was conducted with utmost care to identify all the sources relevant to the themes that emerged. Regarding the study and app use, the Hawthorne effect [49] was considered one of the potential themes in the first analysis, as some users brought up the study setting in the interviews. They were conscious of the ongoing research as they had enrolled in it themselves, and thus, they might have felt a responsibility to use the app. However, it was not possible to determine whether the Hawthorne effect increased or decreased the app use. Therefore, this theme was removed from the analysis.

Declaration of Bias

To avoid researcher bias in the interviews, the interviewees were encouraged to answer the questions frankly and sincerely, and they were assured that there were no right or wrong answers. Interviewers from both teams also tried to avoid any steering of the interviewees in any direction.

Although we cannot be completely certain that the interviewees’ responses fully portrayed their experiences, we trust that they attempted to answer the questions as honestly and sincerely as possible. This trust was further enhanced by the fact that both positive and negative experiences were discussed during the

https://humanfactors.jmir.org/2023/1/e40579
Conclusions

It is important to “know your audience” to predict the potential factors that could hinder use, as it is easier to deal with those factors up front. Some of these factors could be avoided entirely, especially those linked to the design of the system.

Factors associated with the users could be harder to avoid, especially if they are not recognized beforehand. Many of the factors presented in this paper may seem somewhat universal, such as being busy due to work or the bug types found in the system. However, there are other factors that are much harder, perhaps impossible, to counter, such as negative situations in users’ lives (eg, noisy neighbors or the death of a family member).

The PSD postulates present logical aspects and concerns for designing persuasive or other types of systems. However, it may not always be easy to apply them in practice if time and resources are scarce. No one wants to build flawed or buggy systems, but even so, many information systems projects fail.

This is a universal problem, and it comes down to the 3 well-known constraints of the project management triangle and system quality: cost, time, and scope. It is not possible to change only one constraint without affecting quality. Therefore, if the scope, cost, and time are not balanced, it will be challenging to build persuasive (or any other) systems.

Persuasive principles are tools in the design toolbox that can motivate and engage users to strive for behavior change, and in the best-case scenario, they lead to support systems becoming obsolete because users reach their personal goals. However, the persuasive principles are not silver bullets. Careful consideration is required in terms of when and how they should be used. It is crucial to “know the audience,” so the right tools can be selected from the toolbox and put into use to support the users of the designed system in their behavioral change processes.

Implications

Our paper has the following implications:

1. This paper increases knowledge regarding microentrepreneurs, which can be generalized to people with dual roles for example in terms of work and study. Increasing current knowledge about target groups for persuasive design is vital, as studies on this subject are rare. Persuasive design seeks to motivate and engage users to use systems, but limited knowledge about target groups can lead to decreased persuasion. Conversely, increased knowledge could lead to better opportunities to persuade users.

2. Drawing on the PSD model, the paper proposes context-specific solutions to several issues that hinder or reduce the use of similar systems. However, we acknowledge that everything cannot be “designed away.”

3. The paper discusses the role of PSD postulates in improving systems, which has implications for both researchers and designers. Moreover, this paper contributes to the knowledge on how the postulates can be used or aligned for both research and design.

4. This paper also presents a PSD-based solution for a “strict” interpretation (of the autonomy aspect) of SDT in terms of navigation and user freedom. Through the use of personalized tunneling, it should be possible to provide use paths for users based on their own choices, thus not contradicting SDT.

5. We believe that this paper can function as an example of how to use thematic analysis to (1) increase knowledge on target groups through inductive analysis and (2) find theory-based solutions for issues through deductive analysis.

6. We have demonstrated one way to tie inductive thematic analysis with theory commitment, in this case, with persuasion event contexts from the PSD model. Persuasion event contexts are not common, and thus this paper provides an important example of using such an analysis in research.

Future Research

Other target groups with similar issues should be studied in the context of persuasive mHealth apps to uncover similarities or differences between different groups. This would help to generalize the findings regarding persuasive mHealth apps. Future studies could also examine personalized tunneling in terms of app use engagement, which could be helpful for many users.

Acknowledgments

This study is part of Promo@Work research project, which was funded by the Strategic Research Council at the Academy of Finland, contract 303430 (Finnish Institute of Occupational Health) and contract 303431 (University of Oulu, OASIS).

The authors wish to thank all their colleagues from the Promo@Work research consortium.

Conflicts of Interest

All authors were involved in the development of the app with different roles.

Multimedia Appendix 1

Question set 1.

[PDF File (Adobe PDF File), 74 KB - humanfactors_v10i1e40579_app1.pdf]
References


Abbreviations

mHealth: mobile health
PSD: Persuasive Systems Design
SDT: self-determination theory
SME: small and medium-sized company
TTM: transtheoretical model
Barriers to and Facilitators of the Implementation of Digital Mental Health Interventions as Perceived by Primary Care Decision Makers: Content Analysis of Structured Open-Ended Survey Data

Anders Brantnell1,2, PhD; Serdar Temiz3, PhD; Enrico Baraldi3, PhD; Joanne Woodford2, PhD; Louise von Essen2, PhD
1Division of Industrial Engineering and Management, Department of Civil and Industrial Engineering, Uppsala University, Uppsala, Sweden
2Healthcare Sciences and e-Health, Department of Women’s and Children’s Health, Uppsala University, Uppsala, Sweden
3Division of Industrial Engineering and Management, Department of Civil and Industrial Engineering, Uppsala University, Uppsala, Sweden

Corresponding Author:
Anders Brantnell, PhD
Division of Industrial Engineering and Management
Department of Civil and Industrial Engineering
Uppsala University
Box 169
Uppsala, 751 04
Sweden
Phone: 46 729999825
Email: anders.brantnell@angstrom.uu.se

Abstract

Background: Digital mental health represents a way to increase access to evidence-based psychological support. However, the implementation of digital mental health in routine health care practice is limited, with few studies focusing on implementation. Accordingly, there is a need to better understand the barriers to and facilitators of implementing digital mental health. Existing studies have mainly focused on the viewpoints of patients and health professionals. Currently, there are few studies about barriers and facilitators from the perspective of primary care decision makers, that is, the persons responsible for deciding whether a given digital mental health intervention should be implemented in a primary care organization.

Objective: The objectives were to identify and describe barriers to and facilitators of the implementation of digital mental health as perceived by primary care decision makers, evaluate the relative importance of different barriers and facilitators, and compare barriers and facilitators reported by primary care decision makers who have versus have not implemented digital mental health interventions.

Methods: A web-based self-report survey was conducted with primary care decision makers responsible for the implementation of digital mental health in primary care organizations in Sweden. Answers to 2 open-ended questions about barriers and facilitators were analyzed through summative and deductive content analysis.

Results: The survey was completed by 284 primary care decision makers—59 (20.8%) decision makers representing implementers (ie, organizations that offered digital mental health interventions) and 225 (79.2%) respondents representing nonimplementers (ie, organizations that did not offer digital mental health interventions). Overall, 90% (53/59) of the implementers and 98.7% (222/225) of the nonimplementers identified barriers, and 97% (57/59) of the implementers and 93.3% (210/225) of the nonimplementers identified facilitators. Altogether, 29 barriers and 20 facilitators of implementation were identified related to guidelines; patients; health professionals; incentives and resources; capacity for organizational change; and social, political, and legal factors. The most prevalent barriers were related to incentives and resources, whereas the most prevalent facilitators were related to the capacity for organizational change.

Conclusions: A number of barriers and facilitators were identified that could influence the implementation of digital mental health from the perspective of primary care decision makers. Implementers and nonimplementers identified many common barriers and facilitators, but they differ in terms of certain barriers and facilitators. Common and differing barriers and facilitators identified by implementers and nonimplementers may be important to address when planning for the implementation of digital mental health interventions. For instance, financial incentives and disincentives (eg, increased costs) are the most frequently mentioned...
digital mental health; implementation; barriers; facilitators; internet-based cognitive behavioral therapy; survey; decision makers

**Introduction**

**Background**
Common mental health problems, such as depression and anxiety, represent substantial global health challenges [1]. Depression is estimated to be the third-leading cause of disability globally [2], and approximately 29% of all people will be affected by an anxiety disorder during their lifetime [3]. Cognitive behavioral therapy (CBT) delivered face to face is a common and effective treatment for depression and anxiety [4]. However, face-to-face treatments require large organizational resources and visits to health care providers’ offices. Digital mental health represents a way to improve access to care [5] and decrease care costs [6]. Digital mental health can be defined as mental health services and interventions delivered through the internet, telephone, or connected technologies [7]. Internet-administered CBT (ICBT) is a form of digital mental health and has been shown to be as effective as face-to-face CBT for the treatment of depression and anxiety [4]. However, although there is a growing body of research showing the efficacy [8] and cost-effectiveness [9] of ICBT for common mental disorders such as depression and anxiety [10-12], studies of the implementation of ICBT in routine health care practice are limited [13].

To enable implementation and increase access to digital mental health, there is a need to understand aspects that may influence the implementation of digital mental health interventions, that is, barriers and facilitators of implementation. Studies of the implementation of digital mental health are relatively scarce [13], and only a few reviews have identified barriers to and facilitators of implementation [14-18]. By focusing only on the views of patients and health professionals, existing studies have identified barriers such as negative attitudes toward digital mental health [14-16], the lack of suitability of digital mental health for various mental health problems [14-16], low computer literacy [15-17], the lack of training for health professionals [17], and existing infrastructure [17]. Some identified facilitators include training for health professionals [14,16], mild symptoms [14,16], and ease of use [14]. A recent theoretical overview of digital mental health interventions [18] identified barriers such as privacy and security concerns, usability issues from patients’ point of view, patients’ knowledge and skills, and clinicians’ skills and capabilities. One qualitative study exploring mental health professionals’ perspectives about digital mental health implementation identified barriers, such as negative attitudes of clinicians, existing infrastructure, and “one solution does not fit all,” and facilitators, such as the packaging solutions [19]. Furthermore, continued implementation is also a challenge, with a recent review identifying 131 empirical studies of the rapid deployment of digital mental health interventions as a response to the COVID-19 pandemic, with several barriers identified regarding long-term sustainability [20].

Given the few studies focusing on the implementation of digital mental health, it is reasonable to look broadly into digital health implementation. Existing studies in the area have identified several factors that could hinder or facilitate the implementation of interventions. For example, a review studied the factors influencing the adoption of digital applications by health care professionals and identified 101 studies exploring barriers to and facilitators of implementation [21]. Some of the most frequent facilitators of implementation were the usefulness of the innovation and compatibility, whereas some of the most frequent barriers were related to the lack of knowledge among health care professionals and the lack of compatibility [21].

A review of studies (n=16) of the implementation of digital technologies to support patients with amyotrophic lateral sclerosis identified several facilitators of implementation, such as positive attitudes of health care professionals and the training of health care professionals, and barriers, such as negative attitudes of health care professionals and feasibility [22]. Another review of the barriers to the use of digital health by older adults identified 57 studies detailing barriers, with the most frequent barriers being the lack of interest and cost of use [23]. A review focusing on digital health for self-management of hypertension included studies (n=14) that identified barriers to and facilitators of implementation [24]. Some of the most frequent facilitators were access to technology, patient knowledge, and ease of use. In contrast, some of the most frequent barriers were the lack of evidence and added workload [24].

However, none of the digital mental health or digital health reviews identify barriers and facilitators experienced by health care decision makers, that is, the professionals who take the decision to implement or disregard new solutions. Although not included in reviews, there are some qualitative studies that have explored the barriers and facilitators experienced by health care decision makers. A recent qualitative study in Sweden explored policy makers’ views (ie, those who formulate rules and regulations regarding digital health at the regional level, such as politicians) about barriers to and facilitators of the implementation of digital health [25]. Some identified barriers included uncertainty about the impact of digital health on health professionals and the lack of resources for digital health, whereas facilitators included citizens’ preferences and a strong societal push for digital health [25]. Another qualitative study focusing on barriers to the implementation of digital mental health in the United Kingdom, as experienced by health decision makers (health commissioners), identified barriers such as the lack of decision-maker knowledge about the technology, digital literacy among users and decision makers, high risk of investing
in digital mental health, funding issues, and digital interventions not being suitable for all patients [26]. In addition, our previous findings from a web-based cross-sectional quantitative survey about barriers to and facilitators of the implementation of ICBT experienced by primary care decision makers identified a number of barriers to and facilitators of implementation. However, the quantitative survey focused on comparing barriers and facilitators between implementers and nonimplementers but did not capture frequency and thus decision makers’ preferences [27].

**Objectives**

The objectives of this study were (1) to identify and describe barriers to and facilitators of the implementation of digital mental health as perceived by primary care decision makers; (2) to evaluate the relative importance of the barriers and facilitators; and (3) to compare the barriers and facilitators between primary care decision makers who have implemented versus have not implemented digital mental health.

**Methods**

**Study Design**

A web-based self-report survey was conducted between February 2016 and May 2016 with decision makers responsible for the implementation of digital mental health in primary care organizations in Sweden. The survey focused on the implementation of ICBT for depression and anxiety disorders. Answers to the structured open-ended questions in the survey are reported in this paper. Results from the rest of the survey have been reported elsewhere [27].

**Setting**

Sweden was one of the first countries to conduct research on ICBT for depression and anxiety [28]. Swedish national clinical guidelines recommend that CBT and ICBT be provided to adults with mild and moderate levels of depression and anxiety [29]. However, the implementation of ICBT is still in its infancy [30].

Sweden is divided into 21 geographically spread regions that are responsible for health care provision. Each region has several private and public primary care organizations that are publicly funded and thus operate under the same conditions, for instance, in terms of financial resources and adherence to national guidelines. Primary care is the first point of care for patients with mental health problems. The size of the primary care organizations varies in terms of listed patients ranging from 3000 to 30,000. In addition to publicly funded primary care organizations, there are private companies specialized in digital mental health.

Primary care organizations are able to access ICBT through three means: (1) contracting a private company to deliver digital mental health, including support; (2) procuring ICBT program licenses from companies and providing support by themselves; or (3) connecting to the Platform for Support and Care run by the Swedish Association of Local Authorities. Through the Platform for Support and Care, primary care organizations can access ICBT programs developed by private companies or other organizations. There is a cost for connecting to the Platform for Support and Care and for purchasing the treatment programs with or without therapist support. There is also a cost to patients. In the Stockholm Region (one of the 21 regions in Sweden), a web-based meeting with therapist support costs approximately €25 (US $26.7), and costs >€130 (US $138.8) for a patient during the same year will be covered by the public insurance. Furthermore, for patients, it is possible to access ICBT through a private company; for instance, a company charges €75 (US $80.1) for the first meeting and, subsequently, €75 (US $80.1) per week.

As a response to the COVID-19 pandemic, there has been a rapid increase of digital health solutions globally [20]. Available data from 10 regions (including many of the large regions in Sweden) show that 1781 treatments for digital mental health started in 2019 and 4573 started in 2022, when the pandemic had passed, and indicate a modest increase in the provision of digital mental health treatments since the COVID-19 pandemic (for details, refer to Multimedia Appendix 1 [30-32]). Available data do not cover all regions, and according to estimations, 17,800 digital mental health treatments started in 2021 [31]. However, when compared with the number of registered cases of major depression (670,980/5,397,675, 12.43%) and anxiety disorders (536,279/5,397,675, 9.94%) in the first Primary Care Registry in Sweden [33] (common mental health conditions that may be treated with ICBT), the number of digital mental health treatments started still appears to be very low.

**Recruitment and Study Procedures**

Study participants were directors of Swedish primary care organizations. A list of 1156 primary care organizations was compiled, and an invitation was sent to all decision makers. Invitations were initially sent through regular mail and were followed up via telephone and emails. Invitations included a letter explaining the survey and information needed to participate, link to the survey, participation number, and password. Participants who completed the survey provided informed consent through the survey platform (SurveyMonkey [Momentum Global Inc]). No incentives were offered for survey completion. Participants who did not complete the survey within 2 weeks received up to 2 telephone reminders and 1 email. Details about the study procedures are reported elsewhere [27].

**The Survey**

Answers to the 2 open-ended questions in the survey are reported in this study. The following questions were posed:

1. According to your understanding, what are the most important factors that hinder the introduction of ICBT programs? Please indicate a maximum of 5 factors.
2. According to your understanding, what are the most important factors that facilitate introduction of ICBT programs? Please indicate a maximum of 5 factors.

Questions were posed at the end of the survey and were preceded by 37 Likert-scale questions about barriers and facilitators (for details about the survey, refer to the paper by Brantnell et al [27]).
Data Analysis

Data analysis follows summative content analysis, which is a suitable approach to analyze large amounts of open-ended survey data [34]. As there is an abundance of studies of barriers to and facilitators of the implementation of health care interventions, deductive content analysis complemented the summative content analysis [35] guided by the comprehensive integrated checklist of determinants of practice (the Tailored Implementation in Chronic Diseases [TICD] checklist) [36]. The TICD checklist [36] divides barriers and facilitators into each of the seven domains: (1) guidelines; (2) health professionals; (3) patients; (4) professional interaction; (5) incentives and resources; (6) capacity for organizational change; and (7) social, political, and legal factors. The checklist is based on a rigorous review of existing studies of barriers to and facilitators of implementation [36] and provides a good basis for identifying barriers to and facilitators of implementation.

The survey that was reported by Brantnell et al [27] adjusted the TICD checklist according to the Swedish conditions and the Likert-scale approach of the survey questions and thus originated from 5 domains. With open-ended structured data, there was no need to adjust the original TICD checklist because the domains, barriers, and facilitators that would be irrelevant would not be included in the analysis. The analysis was conducted in 6 steps.

First, following a summative content analysis approach [34], data were divided into four small blocks administered through separate Microsoft Excel files: (1) barriers mentioned by implementers (ie, decision makers of organizations that had implemented ICBT); (2) barriers mentioned by nonimplementers (ie, decision makers of organizations that had not implemented ICBT); (3) facilitators mentioned by implementers; and (4) facilitators mentioned by nonimplementers. Second, the Leximancer software was used to identify the most frequent words in each of the 4 data blocks. Subsequently, Excel files were searched for each of the frequent words. To identify possible synonyms for each word, a web-based database, Synonymer.se [37] was used, and the words were added in the search. When applicable, some area-specific synonyms that were not identified by Synonymer.se [37] were added. For example, synonyms to “staff” were “therapist,” “speech therapist,” “psychologist,” “the one treating patients” (behandlare in Swedish), and “medical doctor.”

Third, all hits in the Excel files were marked, and frequent words and phrases were copy-pasted into a Microsoft Word file. The frequency of the copy-pasted words and phrases was recorded. At this stage, no interpretation of data was conducted, but similar words and phrases were combined into large units. Fourth, the words and phrases were translated into English to try to maintain the Swedish phrasing while also capturing the essence of the words and phrases. Fifth, following deductive thematic analysis [35], two researchers independently placed the words and phrases into the TICD checklist under suitable domains (eg, guidelines) and barriers and facilitators (eg, the accessibility of the intervention). Sometimes, the respondents did not provide the direction of the barrier and facilitator. An example of this could be “health professional interest.” In such cases, we modified the word or phrase to include the direction such as “health professionals not interested” (if provided as an answer regarding barriers) or “health professionals interested” (if provided as an answer regarding facilitators). There are some overlaps in the TICD checklist, and thus, the 2 researchers placed some words and phrases under >1 barrier and facilitator. If words and phrases did not fit with existing barriers and facilitators, new barriers and facilitators were created and integrated into the checklist. To increase the credibility of the deductive analysis, coding was conducted by 2 independent coders, which is a recommended procedure when using an existing checklist or framework [38].

An internal workshop was conducted to compare and discuss the outcomes from the deductive analysis. All disagreements were solved through discussion during the workshop. A decision was made to place all words and phrases that lacked a subject (ie, the actor experiencing the barrier or facilitator) such as “leadership” under leadership barriers and facilitators because it was the decision makers who answered the survey. In many cases, respondents provided the subject such as patients or health professionals, and thus, when the subject was missing, a reasonable conclusion was that the words and phrases referred to leadership. While placing words and phrases into the TICD checklist, their frequencies were recorded. All words and phrases mentioned by at least 2 participants were included. Throughout the process, all the authors were involved in discussing and following up on the analysis to increase rigor and trustworthiness [39]. Finally, following the deductive content analysis, the frequencies of each barrier and facilitator were summarized using descriptive statistics, and a comparison between implementers and nonimplementers was conducted. The number of barriers and facilitators was counted for implementers and nonimplementers by adding all words and phrases relating to specific barriers and facilitators.

Ethics Approval

The study was performed in accordance with the Swedish ethical law and the Declaration of Helsinki. The ethical review board of Sweden, Uppsala, approved the study (application number 2015/461).

Results

Respondents

A total of 1156 survey invitations were sent, of which 1130 (97.8%) were shown to be eligible. Noneligible answers that were excluded were duplicate answers (13/26, 50%), bankruptcy or closed down (10/26, 38%), and not a primary care organization (3/26, 12%). A total of 284 decision makers answered the 2 open-ended survey questions.

Characteristics of the Decision Makers

Most decision makers (277/284, 97.5%) were health care center directors or chief executive officers. The 3 most frequent professions of decision makers were nurse (154/284, 54.2%), general practitioner (59/284, 20.8%), and physiotherapist (20/284, 7%). Among the respondent organizations, 20.8% (59/284) provided ICBT and were thus implementers. Overall, 63% (179/284) of the decision makers represented public
organizations and the remaining represented private organizations. Both types of organizations were publicly funded.

**Barriers to and Facilitators of Implementation of ICBT**

**Overview**

Altogether, 59 implementers responded, of which 57 (97%) listed the facilitators of implementation and 53 (90%) listed the barriers to implementation. In contrast, 225 nonimplementers responded, of which 210 (93.3%) listed the facilitators of implementation, whereas 222 (98.7%) listed the barriers to implementation.

In total, 29 barriers to and 20 facilitators of the implementation of ICBT were identified (Tables 1 and 2), and these were grouped within 6 domains based on the TICD checklist (guidelines; health professionals; patients; incentives and resources; capacity for organization change; and social, political, and legal factors). No barriers and facilitators were mentioned regarding the seventh domain in the TICD checklist, namely, professional interaction. For detailed outcomes of the summative and deductive content analysis, refer to Multimedia Appendix 2. All the barriers and facilitators are presented in Tables 1 and 2. The most frequently mentioned barriers were related to incentives and resources (ie, availability of necessary resources; 14/53, 26%) and capacity for organizational change (ie, capable leadership; 14/53, 26%; Table 1), whereas the most frequently mentioned facilitators were related to capacity for organizational change (ie, assistance for organizational change; 57/210, 27.1%; Table 2).
Table 1. Barriers mentioned by implementers and nonimplementers, distributed according to the Tailored Implementation in Chronic Diseases checklist.

<table>
<thead>
<tr>
<th>Domains and barriers</th>
<th>Implementers (n=53), n (%)</th>
<th>Nonimplementers (n=222), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Guidelines</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Compatibility</td>
<td>8 (15.1)</td>
<td>_a</td>
</tr>
<tr>
<td>Feasibility</td>
<td>_</td>
<td>23 (10.4)</td>
</tr>
<tr>
<td>Strength of the recommendation</td>
<td>_</td>
<td>20 (9)</td>
</tr>
<tr>
<td>Accessibility of the intervention</td>
<td>2 (3.8)</td>
<td>4 (1.8)</td>
</tr>
<tr>
<td>Quality of evidence supporting the</td>
<td>2 (3.8)</td>
<td>2 (0.9)</td>
</tr>
<tr>
<td>recommendation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effort</td>
<td>_</td>
<td>4 (1.8)</td>
</tr>
<tr>
<td>Clarity</td>
<td>_</td>
<td>2 (0.9)</td>
</tr>
<tr>
<td>Cultural appropriateness</td>
<td>_</td>
<td>2 (0.9)</td>
</tr>
<tr>
<td>Trialability</td>
<td>_</td>
<td>2 (0.9)</td>
</tr>
<tr>
<td><strong>Health professionals</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intention and motivation</td>
<td>10 (18.9)</td>
<td>23 (10.4)</td>
</tr>
<tr>
<td>Nature of the behavior</td>
<td>6 (11.3)</td>
<td>2 (0.9)</td>
</tr>
<tr>
<td>Attitudes</td>
<td>_</td>
<td>14 (6.3)</td>
</tr>
<tr>
<td>Skills needed to adhere</td>
<td>2 (3.8)</td>
<td>2 (0.9)</td>
</tr>
<tr>
<td>Awareness and familiarity with the</td>
<td>_</td>
<td>5 (2.2)</td>
</tr>
<tr>
<td>recommendation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>** Patients**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient motivation and interest</td>
<td>4 (7.5)</td>
<td>15 (6.8)</td>
</tr>
<tr>
<td>Patient behavior</td>
<td>3 (5.7)</td>
<td>5 (2.2)</td>
</tr>
<tr>
<td>Patient preferences</td>
<td>2 (3.8)</td>
<td>11 (4.9)</td>
</tr>
<tr>
<td>Patient beliefs and knowledge</td>
<td>_</td>
<td>4 (1.8)</td>
</tr>
<tr>
<td><strong>Incentive and resources</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Availability of necessary resources</td>
<td>14 (26.4)</td>
<td>29 (13.1)</td>
</tr>
<tr>
<td>Financial incentives and disincentives</td>
<td>5 (9.4)</td>
<td>33 (14.9)</td>
</tr>
<tr>
<td>Information system</td>
<td>3 (5.7)</td>
<td>15 (6.8)</td>
</tr>
<tr>
<td>Availability of supporting infrastructure</td>
<td>3 (5.7)</td>
<td>11 (4.9)</td>
</tr>
<tr>
<td><strong>Capacity for organizational change</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Capable leadership</td>
<td>14 (26.4)</td>
<td>53 (23.9)</td>
</tr>
<tr>
<td>Organizational readiness</td>
<td>6 (11.3)</td>
<td>25 (11.3)</td>
</tr>
<tr>
<td>Assistance for organizational change</td>
<td>5 (9.4)</td>
<td>11 (4.9)</td>
</tr>
<tr>
<td>Mandate, authority, and accountability</td>
<td>2 (3.8)</td>
<td>_</td>
</tr>
<tr>
<td>Regulations, rules, and policies</td>
<td>_</td>
<td>2 (0.9)</td>
</tr>
<tr>
<td><strong>Social, political, and legal factors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health care system</td>
<td>_</td>
<td>18 (8.1)</td>
</tr>
<tr>
<td>Contracts</td>
<td>_</td>
<td>3 (1.4)</td>
</tr>
</tbody>
</table>

*aNot available.*
Table 2. Facilitators mentioned by implementers and nonimplementers, distributed according to the Tailored Implementation in Chronic Diseases checklist.

<table>
<thead>
<tr>
<th>Domains and facilitators</th>
<th>Implementers (n=57), n (%)</th>
<th>Nonimplementers (n=210), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Guidelines</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accessibility of the intervention</td>
<td>8 (14)</td>
<td>36 (17.1)</td>
</tr>
<tr>
<td>Feasibility</td>
<td>5 (8.8)</td>
<td>14 (6.6)</td>
</tr>
<tr>
<td>Strength of the recommendation</td>
<td>4 (7)</td>
<td>2 (1)</td>
</tr>
<tr>
<td>Observability</td>
<td>3 (5.3)</td>
<td>—</td>
</tr>
<tr>
<td>Clarity</td>
<td>3 (5.3)</td>
<td>—</td>
</tr>
<tr>
<td>Compatibility</td>
<td>—</td>
<td>3 (1.4)</td>
</tr>
<tr>
<td><strong>Health professionals</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intention and motivation</td>
<td>11 (19.3)</td>
<td>24 (11.4)</td>
</tr>
<tr>
<td><strong>Patients</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient motivation and interest</td>
<td>4 (7)</td>
<td>5 (2.4)</td>
</tr>
<tr>
<td>Patient beliefs and knowledge</td>
<td>2 (3.5)</td>
<td>6 (2.9)</td>
</tr>
<tr>
<td>Patient behavior</td>
<td>—</td>
<td>6 (2.9)</td>
</tr>
<tr>
<td><strong>Incentives and resources</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Financial incentives and disincentives</td>
<td>2 (3.5)</td>
<td>37 (17.6)</td>
</tr>
<tr>
<td>Availability of necessary resources</td>
<td>5 (8.8)</td>
<td>9 (4.3)</td>
</tr>
<tr>
<td>Information system (people, platform, and technology combined)</td>
<td>3 (5.3)</td>
<td>12 (5.7)</td>
</tr>
<tr>
<td>Availability of supporting infrastructure</td>
<td>2 (3.5)</td>
<td>2 (1)</td>
</tr>
<tr>
<td>Nonfinancial incentives and disincentives</td>
<td>2 (3.5)</td>
<td>2 (1)</td>
</tr>
<tr>
<td><strong>Capacity for organizational change</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assistance for organizational change</td>
<td>11 (19.3)</td>
<td>57 (27.1)</td>
</tr>
<tr>
<td>Capable leadership</td>
<td>5 (8.8)</td>
<td>44 (21)</td>
</tr>
<tr>
<td>Relative strength of supporters and opponents</td>
<td>2 (3.5)</td>
<td>7 (3.3)</td>
</tr>
<tr>
<td>Mandate, authority, and accountability</td>
<td>—</td>
<td>2 (1)</td>
</tr>
<tr>
<td><strong>Social, political, and legal factors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health care system</td>
<td>—</td>
<td>2 (1)</td>
</tr>
</tbody>
</table>

^Not available.

**Guidelines**

Overall, 9 barriers and 6 facilitators were identified regarding the guidelines for ICBT interventions. Implementers most often mentioned the lack of compatibility with existing technology (8/53, 15%) as a barrier. In contrast, nonimplementers most often mentioned the feasibility of the intervention (ie, the extent to which the intervention is practical; 23/222, 10.4%). Of the 9 barriers identified, 2 (the accessibility of the intervention and quality of evidence supporting the recommendation) were mentioned by both implementers and nonimplementers. Both implementers (8/57, 14%) and nonimplementers (36/210, 17.1%) most often mentioned the accessibility of the intervention as a facilitator. Of the 6 facilitators identified, 3 (the accessibility of the intervention, feasibility, and strength of recommendation) were mentioned by both implementers and nonimplementers.

**Health Professionals**

Overall, 5 barriers and 1 facilitator were identified related to health professionals. Implementers (10/53, 19%) and nonimplementers (23/222, 10.4%) most often mentioned the lack of intention and motivation as barriers. Of the 5 barriers identified, 3 (intention and motivation, the nature of the behavior, and skills needed to adhere) were mentioned by both implementers and nonimplementers. Both implementers and nonimplementers mentioned only 1 facilitator—intention and motivation (11/57, 19% for implementers and 24/210, 11.4% for nonimplementers).

**Patients**

Overall, 4 barriers and 3 facilitators were identified related to patients. Implementers (4/53, 8%) and nonimplementers (15/222, 6.8%) most often mentioned the barrier, lack of patient motivation and interest. Of the 4 barriers identified, 3 (patient motivation and interest, patient behavior, and patient
preferences) were mentioned by both implementers and nonimplementers. The most frequently mentioned facilitator by implementers was patient motivation and interest (4/57, 7%). In contrast, the most frequently mentioned facilitator by nonimplementers was patient beliefs and knowledge (6/210, 2.9%) and patient behavior (6/210, 2.9%). Of the 3 facilitators identified, 2 (patient motivation and interest and patient beliefs and knowledge) were mentioned by both implementers and nonimplementers.

**Incentives and Resources**

Overall, 4 barriers and 5 facilitators were identified regarding incentives and resources. Implementers most often mentioned the barrier, the availability of necessary resources (14/53, 26%). In contrast, nonimplementers most often mentioned the barriers, financial incentives and disincentives (33/222, 14.9%) and the information system (33/222, 14.9%). All the 4 barriers identified were mentioned by both implementers and nonimplementers. Implementers most often mentioned the facilitator, the availability of necessary resources (5/57, 9%), whereas nonimplementers most often mentioned the facilitator, financial incentives and disincentives (37/210, 17.6%). All the 5 facilitators identified were mentioned by both implementers and nonimplementers.

**Capacity for Organizational Change**

Overall, 6 barriers and 5 facilitators were identified regarding capacity for organizational change. Both implementers and nonimplementers most often mentioned the barrier, capable leadership (ie, leadership interest and knowledge; 14/53, 26% for implementers and 53/222, 23.9% for nonimplementers). Of the 5 barriers identified, 3 (capable leadership, organizational readiness, and assistance for organizational change) were mentioned by both implementers and nonimplementers. Organizational readiness was not part of the TICD checklist but originated from the summative content analysis and was added to the checklist. Implementers (11/57, 19%) and nonimplementers (57/210, 27.1%) most often mentioned the facilitator, assistance for organizational change. Of the 4 facilitators identified, 3 (assistance for organizational change, capable leadership, and relative strength of supporters and opponents) were mentioned by both implementers and nonimplementers.

**Social, Political, and Legal Factors**

Overall, 2 barriers and 1 facilitator were mentioned by nonimplementers related to social, political, and legal factors. The barrier that was most often mentioned was the health care system (18/210, 8.6%). The health care system was not part of the TICD checklist but was added based on the summative content analysis. The only facilitator was the health care system (2/210, 0.9%).

**Discussion**

**Principal Findings**

A total of 284 decision makers participated in the survey and provided answers to 2 open-ended questions. The majority of respondents (277/284, 97.5%) were health care center directors or chief executive officers. The 3 most common professions among the decision makers were nurses (154/284, 54.2%), general practitioners (59/284, 20.8%), and physiotherapists (20/284, 7%). Out of all the organizations represented, 20.8% (59/284) offered ICBT and were labeled as implementers. Among the implementers, 90% (53/59) identified barriers to implementation, while 97% (57/59) listed facilitators of implementation. On the other hand, among the nonimplementers, 98.7% (222/225) listed barriers to implementation and 93.3% (210/225) listed facilitators of implementation. In total, 29 barriers to and 20 facilitators of implementing ICBT were identified.

Findings identified barriers to and facilitators of the implementation of digital mental health related to 6 domains in the TICD checklist: guidelines; health professionals; patients; incentives and resources; capacity for organizational change; and social, political, and legal factors. First, we conducted summative content analysis based on the responses. During this phase, we were able to capture barriers and facilitators, as expressed by the respondents. Second, we connected the responses with the TICD checklist [36]. No barriers or facilitators were identified related to the TICD checklist domain, professional interaction. In addition, we identified 3 new barriers and facilitators that were added to the TICD checklist: the availability of supporting infrastructure (domain: incentives and resources), organizational readiness (domain: capacity for organizational change), and the health care system (domain: social, political, and legal factors).

Findings show that the most frequently mentioned barriers related to availability of necessary resources (14/53, 26%) and capable leadership (14/53, 26%), whereas the most frequently mentioned facilitators related to assistance for organizational change (57/210, 27.1%). Existing studies of the implementation of digital mental health [26] and digital health [25] interventions focusing on decision makers imply that the availability of necessary resources is an important barrier to implementation. This barrier (availability of necessary resources) is further supported by existing studies of policy makers’ use of evidence [39] and barriers to implementation related to third-sector actors providing health care [40]. Our findings align with a review focusing on health professionals’ views that identified important barriers and facilitators related to organizations, systems, and health professionals including assistance for organizational change [16]. However, our findings also suggest that assistance for organizational change also relates to decision makers. Education and support, which are important components of organizational change (for details, refer to the codes in Multimedia Appendix 2—under the column heading Words and phrases mentioned by respondents, grouped as Assistance for organizational change), have been identified as barriers in existing studies focusing on implementation facilitators for third-sector actors providing health care [40] and studies of health policy makers’ use of evidence [41].

Assistance for organizational change and capable leadership are closely related, and thus, it is unsurprising that capable leadership is one of the most frequently mentioned barriers to implementation. That is, if managers are not trained and educated, they will not be able to support implementation.
Increasing knowledge requires an implementation strategy [42], and in general, implementation strategies are reported to have between 4% and 10% effect sizes in changing behavior [43]. Furthermore, the lack of knowledge could be dependent on other factors such as attitudes and outcome expectations, and thus, addressing these barriers is also needed [42]. Therefore, is it not likely that COVID-19 and similar disruptions could wipe out complex and sizable barriers to implementation, and thus, a reasonable assumption is that many barriers are persistent and require structured implementation efforts rather than sudden external pressure. These assumptions are also supported in existing studies of barriers to sustain digital mental health interventions after the COVID-19 pandemic [20].

Comparing our findings with those of existing studies of frequent barriers and facilitators, we underscore 3 important findings. First, existing studies of digital mental health have identified very few barriers and facilitators related to guidelines or the therapy in itself, such as the strength of the recommendation (ie, 1 solution does not fit all) [19,26]. However, the frequently mentioned guideline-related barriers and facilitators in our findings, the lack of compatibility [21], feasibility [22,44], and the accessibility of the intervention [24], are well established in existing studies of digital health implementation. Second, existing studies of digital mental health [26] and digital health [25] with focus on decision makers do not raise the importance of health professionals' intention and motivation as both barriers to and facilitators of implementation. In our findings, this barrier and facilitator was the most frequently mentioned related to health professionals, and it is also identified in existing studies of digital mental health [19] and digital health [22]. Third, patient’s motivation and interest are identified as important barriers to and facilitators of implementation in existing studies of digital health implementation [23,24], but they are not very prevalent among our findings regarding the implementation of digital mental health from the perspective of decision makers.

Implementers and nonimplementers identified a number of similar barriers and facilitators relating to 4 (health professionals, patients, incentives and resources, and capacity for organizational change) of the 6 domains. Most similarities were identified in relation to incentives and resources. However, there were differences in how frequently these barriers were mentioned. Implementers report availability of necessary resources as the most frequent barrier and facilitator, whereas these are not the most frequently reported by nonimplementers. These findings imply that the implementation of digital mental health interventions is not dependent on available resources, albeit may be hindered by lack of them. This, in turn, could be encouraging for nonimplementers that lack the necessary resources to invest in digital mental health. One way to facilitate implementation could be to communicate to nonimplementers, especially persons with budgetary responsibilities, that maintaining implementation requires additional resources. Whether maintaining implementation requires extra resources is an empirical question for further studies.

Financial incentives and disincentives are the barriers and facilitators most frequently mentioned by nonimplementers but not by implementers, which implies that nonimplementers perceive structural hinder for implementation related to financial aspects such as the reimbursement system and increased costs. In contrast, implementers do not perceive the financial incentives and disincentives as highly problematic, which, in turn, could be motivating for nonimplementers that assume these to be sizable barriers to implementation. One way to facilitate implementation could be to provide more information to nonimplementers regarding the actual costs related to the implementation of digital mental health. Whether the benefits of digital mental health interventions are related to financial incentives or other aspects such as improved care warrants further studies.

The most obvious difference between implementers and nonimplementers was found in barriers and facilitators related to guidelines. The 2 barriers most frequently mentioned by nonimplementers related to guidelines are the feasibility and strength of recommendation, whereas these are not mentioned by implementers. Whether these 2 barriers are real barriers based on experience or only based on assumptions is unclear; however, overcoming these barriers, for instance, through education could improve the possibilities for implementation, and thus, it could be beneficial to educate nonimplementers regarding the feasibility and strength of the recommendation. Implementers most frequently mention compatibility as a barrier related to guidelines, whereas nonimplementers do not mention this barrier, which implies that implementers perceive that there is not an optimal fit between the digital mental health intervention and existing work practices. This type of barrier is difficult to overcome because it is at the core of the intervention, that is, starting to use the intervention requires work with computers and thus requires a more complex implementation strategy targeting possible barriers such as digital literacy, attitudes toward digital mental health, and adaptation of existing work routines to accommodate the provision of digital mental health.

**Limitations**

Our study has some limitations. First, we collected structured, open-ended data using a survey, which is not an optimal way to gain an in-depth understanding of the barriers and facilitators because no follow-up questions can be posed. However, we followed a well-structured and rigorous analysis process that should be able to provide a good overview of barriers to and facilitators of ICBT implementation in Sweden from the perspective of decision makers. Second, we collected data from 1 country. Sweden has publicly funded health care, with good access to care. There could be certain contextual differences between different health care systems, but some barriers and facilitators such as capable leadership could apply to several contexts and digital mental health more generally. Whether capable leadership and the other identified barriers and facilitators also apply to other health care systems and technologies is an empirical question for further studies.

Third, our data were collected in 2016. However, these data are still relevant for several reasons: neither the intervention nor the context has changed substantially since 2016. Swedish primary care organizations can still access ICBT in several ways, and it is often the primary care director who can make the decision regarding whether to offer ICBT.
that during the past years, the technology has probably matured and could have become less costly and more easily available. However, it is unlikely that disruptions such as COVID-19 and technology advancements have overcome the complex and sizable barriers to implementation that were identified, and thus, a reasonable assumption is that many barriers are persistent and require structured implementation efforts. For instance, it is unlikely that there has been substantial increase in available resources for digital mental health, and although regions in Sweden are investing in digital innovations, mental health has not been their priority [45]. Similarly, leadership and public health workforce capacity building requires structured and complex implementation efforts [46].

Moreover, despite expectations of a massive increase of digital mental health treatments after the COVID-19 pandemic, the number of treatments in Sweden was still modest during the COVID-19 pandemic (refer to the Methods section), which makes the identified barriers and facilitators relevant. One reason for the relatively low numbers of digital mental health treatments provided during the COVID-19 pandemic could be the Swedish government’s decision to not use lockdown measures, meaning that health care was still provided face to face [47,48]. However, we do not present any data about digital mental health provision in other comparable countries, such as Denmark and Norway, which have similar public health systems but adopted different COVID-19 responses [48]. Thus, how and whether COVID-19 influenced digital mental health provision in these countries is an empirical question for further studies.

Fourth, the 2 open-ended questions were posed at the end of the survey, after asking 37 questions to be answered on a Likert scale, which could risk priming the responses. However, we deem this risk to be low because the abundance of Likert-scale questions would rather provide information overload than clear advice about possible barriers and facilitators for a respondent who has not considered them beforehand.

Conclusions
Globally, the COVID-19 pandemic resulted in rapid deployment of various digital health solutions such as telehealth and videoconferencing to provide continued care despite distancing requirements. However, given the complex nature of digital mental health solution implementation, it is not probable that implementation based on sudden external pressure will be maintained. So far, few studies have examined the barriers to and facilitators of the implementation of digital mental health, and even fewer studies have examined the perspectives of decision makers. In this study, we report about various barriers and facilitators related to guidelines; health professionals; patients; incentives and resources; capacity for organization change; and social, political, and legal factors. Commonly reported barriers, by both implementers and nonimplementers, related to incentives and resources, whereas common facilitators were related to capacity for organizational change, and most differences were identified in relation to guidelines. Understanding similarities and differences can provide advice to future implementers of digital health regarding barriers and facilitators to take into consideration and inform the development of implementation strategies.

Acknowledgments
The authors thank all decision makers who participated in the study. The authors thank several colleagues for their input. This study was supported by Uppsala University Psychosocial Care Programme (U-CARE), which is a strategic research environment funded by the Swedish Research Council (dnr 2009–1093). The funder did not contribute to the design or conduct of the study and did not participate in the analysis of the results.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Started digital mental health treatments during 2018 and 2022. The file illustrates the number of started digital mental health treatments during 2018 and 2022. Data in the file are based on these studies [30,31,33].

[DOCX File, 110 KB - humanfactors_v10i1e44688_app1.docx ]

Multimedia Appendix 2
Outcomes of summative content analysis and deductive analysis based on the Tailored Implementation in Chronic Diseases checklist. The file describes in detail the outcomes of the 2 coding procedures, listing the words and phrases, barriers, facilitators, and domains.

[DOCX File, 28 KB - humanfactors_v10i1e44688_app2.docx ]

References


Abbreviations

CBT: cognitive behavioral therapy
ICBT: internet-administered cognitive behavioral therapy
TICD: Tailored Implementation in Chronic Diseases
Factors Affecting Digital Tool Use in Client Interaction According to Mental Health Professionals: Interview Study

Lauri Lukka¹, MA; Veli-Matti Karhulahti², PhD; J Matias Palva¹,³,⁴, PhD
¹Department of Neuroscience and Biomedical Engineering, Aalto University, Espoo, Finland
²Faculty of Humanities and Social Sciences, University of Jyväskylä, Jyväskylä, Finland
³Neuroscience Center, Helsinki Institute of Life Science, University of Helsinki, Helsinki, Finland
⁴Centre for Cognitive Neuroimaging, Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, United Kingdom

Corresponding Author:
Lauri Lukka, MA
Department of Neuroscience and Biomedical Engineering
Aalto University
Rakentajanaukio 2
Espoo, 02150
Finland
Phone: 358 440375666
Email: lauri.lukka@aalto.fi

Abstract

Background: Digital tools and interventions are being increasingly developed in response to the growing mental health crisis, and mental health professionals (MHPs) considerably influence their adoption in client practice. However, how MHPs use digital tools in client interaction is yet to be sufficiently understood, which poses challenges to their design, development, and implementation.

Objective: This study aimed to create a contextual understanding of how MHPs use different digital tools in clinical client practice and what characterizes the use across tools.

Methods: A total of 19 Finnish MHPs participated in semistructured interviews, and the data were transcribed, coded, and inductively analyzed.

Results: We found that MHP digital tool use was characterized by 3 distinct functions: communication, diagnosis and evaluation, and facilitating therapeutic change. The functions were addressed using analog tools, digitized tools that mimic their analog counterparts, and digital tools that use the possibilities native to digital. The MHP-client communication included various media alongside face-to-face meetings, the MHPs increasingly used digitized tools in client evaluation, and the MHPs actively used digitized materials to facilitate therapeutic change. MHP tool use was generally characterized by adaptability—it was negotiated in client interactions. However, there was considerable variance in the breadth of MHPs’ digital toolbox. The existing clinical practices emphasized MHP-client interaction and invited incremental rather than radical developments, which challenged the achievement of the scalability benefits expected from digital tools.

Conclusions: MHPs use digitized and digital tools in client practice. Our results contribute to the user-centered research, development, and implementation of new digital solutions in mental health care by classifying them according to their function and medium and describing how MHPs use and do not use them.

(JMIR Hum Factors 2023;10:e44681) doi:10.2196/44681

KEYWORDS
clinical practice; digital mental health interventions; intervention design; mental health applications; mental health professionals; teletherapy; mobile phone
Introduction

Background
Mental disorders are the leading cause of disease burden worldwide [1]. However, a substantial number of people with mental disorders fail to receive adequate support and treatment for their challenges [2]. It is believed that digital technologies can increase the effectiveness, accessibility, and cost-effectiveness of existing treatments, which is a considerable motivation for their development [3-7].

Mental health professionals (MHPs) play a considerable role in how mental disorders are treated and which digital tools and materials are used in clinical practice. They act as gatekeepers for web-based therapies [8] and exercise their influence by recommending digital materials, platforms, and treatments to their clients [9]. Thus, alongside their clients, MHPs constitute a second key user group [10] whose attitudes and needs are vital to understand when designing and developing new digital tools.

MHP and client needs relate through the so-called therapeutic alliance. Bordin [11] describes that it consists of three factors: (1) a positive attachment bond between the MHP and their client, (2) their shared agreement on the therapy goals, and (3) pursuing these goals through tasks in therapeutic interaction. The therapeutic relationship has been found to substantially contribute to the effectiveness of therapy [12,13]. Today, digital tools present changes and possibilities for the therapeutic alliance [14-16]—telehealth solutions facilitate the contact between the MHP and their client, and in counseling, the MHP may facilitate the change through complementary digitized materials, digital therapies, and mobile apps. The therapeutic alliance is becoming digitally enhanced.

MHP Digital Tool Use in Client Practice

Telehealth Solutions
The COVID-19 crisis has substantially changed how MHPs interact with their clients. Before the crisis, only a minority of the interactions occurred remotely. A 2018 published survey found that 57% of US psychologists did not engage in telecounseling, with only 6% delivering >6 hours of telecounseling per week [17]. Landlines and mobile phones were considerably more common. In Portugal, Mendes-Santos et al [9] had similar findings: 30% of psychologists used digital technology to support their clients, most often via telephone, email, and SMS text messages; only 9% used videoconferencing. Such low use likely reflects the numerous barriers to using telehealth solutions, including the perceived dehumanization of the therapeutic environment, client and clinician suitability factors, and the prohibitive costs of the solutions, as well as issues with reimbursement, confidentiality, and data protection [18,19]. However, the pandemic has forced therapists to offer their services remotely [20-22], and today, digital media increasingly facilitate client contact.

Digital Materials
Clients can use psychoeducational materials independently to alleviate their psychiatric symptoms, such as depression [23]. MHPs also often recommend complementary web-based materials to be accessed between therapy sessions [9,24], including websites, forums, blogs, social media, and support groups. The need for recommendations exists as not all abundant web-based materials comply with and reflect the best treatment practices [3]. To help both MHPs and their clients navigate web-based materials, digital mental health hubs have been created. Canadian eMentalHealth.ca, for example, provides information on mental health, self-assessment forms, and contact points to health care, and the 2 million annual users generally find the service positive [25]. In Finland, MentalHub (in Finnish, “Mielenterveystalo”), developed by the Helsinki University Hospital (HUS), serves a similar purpose. It provides psychoeducation, self-guided treatments, symptom navigators, service directories, and internet therapies [26].

Digital Mental Health Interventions
Comprehensive and structured digital treatment programs have also been developed. These emerging therapies have been called computer-assisted therapy [27] and internet-delivered psychological treatments [28]; we use the concept of digital mental health interventions (DMHIs) [29]. A recent meta-analysis found that therapist-supported internet-based interventions yield similar effects as face-to-face cognitive behavioral therapy (CBT) [30]. Indeed, MHP contact is beneficial in motivating and engaging the client in the digital intervention. DMHIs with therapist support are more effective than without it [31,32]. Similar to traditional face-to-face therapy, the quality of the mediated therapeutic alliance during the DMHI contributes to treatment outcomes [33]. Moreover, it appears that the richness of the contact facilitates treatment results—face-to-face support is more effective than telephone support, which is more effective than email support [27].

Unguided, guided, and blended therapies have been actively developed and used. An example of a guided DMHI is the HUS-provided, physician-referred 12-session CBT program for generalized anxiety disorder [34]. The program is theoretically based on several models of anxiety, trained therapists offer support through the program using asynchronous messages, and the program includes persuasive elements such as simulation and reminders to increase client engagement. HUS has developed internet therapies for other mental disorders as well, such as depression, attention-deficit/hyperactivity disorder, and bipolar disorder [35], which are referred to in this research as “national DMHI.” Blended therapies combine both digital and face-to-face interaction [36]. In the United Kingdom, Stawarz et al [37] reported a blended approach comprising initial face-to-face meetings, subsequent web-based therapy sessions, and independent work by the client between sessions. Generally, MHPs are more favorable toward blended approaches than unguided therapies [8,38-40]. This reflects their profession that emphasizes the importance of the healing therapeutic relationship which frames discussions on digital tools.

Mental Health Apps
Finally, the rapid proliferation of smartphones has brought health care to the clients’ pockets, and the use of mental health apps is growing [41]. In contrast to the DMHIs that are often developed and delivered in association with health care organizations and may also be clinician-prescribed and
reimbursed, mental health apps are often distributed directly to consumers and may be used independently of health care contact [42]. In some countries, 50% of mental health service—using youth [43] and 10% of outpatient psychiatry clinic patients have used mental health apps [44]. An example of a commercial mental health app is “Calm,” which offers a mindfulness meditation intervention that can reduce stress [45] and increase well-being [46]. An example of a Finnish publicly funded mental health app is “Chillaa,” which is targeted to youth aged 13 to 15 years and aims to reduce stress and social anxiety [47].

Mental health apps may complement therapies [48], and some MHPs recommend them to their clients. In Portugal, 28% of psychologists had recommended apps to their clients [9], whereas in examining mental services for the youth, Bell et al [43] found that 84% of clinicians had recommended apps to their clients. However, there have been considerable concerns regarding the quality and evidence base of the app content [49-53] as well as their privacy [54], which, together with the lack of guidance [55], diminish their credibility and slow their adoption in mental health care [36].

In summary, MHPs use various digital tools in their client practice. Moreover, the use of fully digital and blended DMHIs is growing—a trend that is occurring alongside the proliferation of mobile mental health apps. The growing adoption of digital technologies in society and health care and the specific changes in the digital mental health landscape frame MHP attitudes regarding digital tools.

Study Aims

Previous research has examined MHP adoption of digital technologies, their attitudes toward them, and the factors influencing their implementation. However, less attention has been paid to how MHPs use digital tools in client practice. We posit that a qualitative, user-centered approach can provide rich, in-depth insights into the MHP working context and their attitudes, needs, preferences, and behavior regarding digital tools [57-60]—factors vital to their design, development, and implementation.

We followed the hypothesis-generating qualitative research tradition [61] and focused on how MHPs describe their client practice [62]. How we conceptualized qualitative analysis in this study reflected a perspective aptly described by Fossey et al [63]:

*Qualitative research aims to address questions concerned with developing an understanding of the meaning and experience dimensions of humans’ lives and social worlds. Central to good qualitative research is whether the research participants’ subjective meanings, actions, and social contexts, as understood by them, are illuminated.*

Initially, the study was planned to better understand MHP views and needs regarding DMHIs with game elements. The aim was broken down into 2 areas the MHPs were familiar with: how they perceived and used digital tools and how they viewed digital games in client practice. In the very first interviews, it was discovered that these 2 areas were separate—digital tool use was MHP initiated, whereas playing digital games addressed client behavior. Thus, this study focused on the first area, which was further broken down into two specific research questions (RQs): How do MHPs use different digital tools in client practice? (RQ 1) and What characterizes MHPs’ digital tool use in client practice in general? (RQ 2).

To further the study aims, 19 semistructured interviews with Finnish MHPs were conducted, analyzed inductively, and reported under the 2 RQs.

Methods

The study was conducted in 3 phases: recruitment, interview, and analysis. A survey was used to gather background information, whereafter MHPs were invited to a semistructured interview, and the transcribed interview data were analyzed inductively.

Ethics Approval

The study was approved by the Aalto University research ethics committee (D/508/03.04/2022), and the research design was preregistered in the Open Science Framework [64].

Data and Sampling

The guiding principle in participant recruitment was maximum local variation—the recruitment aimed to gather a diverse sample of Finnish MHPs with various educational backgrounds, who worked in different organizational contexts, and with various client populations in health care. Finnish health care is primarily public and organized by municipalities [65] and tiered into low-threshold basic-level services for those with less severe disorders and specialized psychiatric services for clients with more severe disorders. The Finnish Student Health Service provides mental health care services for university students [66], and occupational health care provides health care services and brief counseling for the workforce. Rehabilitative psychotherapy delivered by licensed psychotherapists can be reimbursed for up to 3 years [67].

The study participants were recruited through social media, local professional association channels, and health care organizations. Snowballing was used to recruit professionals from the expert networks of the interviewees. The participant recruitment advertisements highlighted that the interviewees were not expected to have experience using digital tools and therapies to welcome participants with various levels of experience. The study inclusion criteria were (1) being a licensed health care professional, (2) working with mental health, and (3) having at least one customer weekly. Interviews were conducted with a Finnish interview frame; thus, non—Finnish-speaking participants were not included. The recruitment and interviews took place between May 11, 2022, and September 8, 2022.

A recruitment link shared on the web led the possible participant to a web-based survey in Finnish created using Webropol software (Webropol Limited) and included an informed consent form and privacy notice. According to the service use statistics, 833 people opened the digital questionnaire, of whom 109 (13.1%) began to answer it, with 80 (73.4%) of the 109 MHPs completing it. A total of 34 respondents indicated their
willingness to participate in the interviews, of whom 24 (71%) were contacted by the first author via email. In total, 9% (3/34) of the respondents did not respond to the inquiry, and 6% (2/34) withdrew before the interview: one because of a lack of time and another because of their considerable prejudices against the topic.

The concept of saturation [68] was used to evaluate the sufficiency of the sample. The first author evaluated saturation using analytic memoing conducted after each interview [69] and during coding, which was carried out in parallel to the interviews. After approximately 12 interviews, the first author found that they contributed less and less new information. After a collaborative reflection of the analytic memos and initial codes with the second author, it was deemed that the data from 19 interviews were sufficient to answer the RQs. Therefore, not all 34 respondents willing to be interviewed were contacted. The sample size of 19 is aligned with a recent systematic review that found that 9 to 17 interviews reached saturation in homogenous study populations such as ours [70], as well as previous research on interview sample sizes [71]. The characteristics of the interviewees are described in Table 1 and individually in Table 2.

Table 1. Characteristics of the interviewees (n=19).

<table>
<thead>
<tr>
<th>Variable and category</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender, n (%)</td>
<td></td>
</tr>
<tr>
<td>Woman</td>
<td>13 (68)</td>
</tr>
<tr>
<td>Man</td>
<td>6 (32)</td>
</tr>
<tr>
<td>Age range (years), n (%)</td>
<td></td>
</tr>
<tr>
<td>18-29</td>
<td>1 (5)</td>
</tr>
<tr>
<td>30-39</td>
<td>3 (16)</td>
</tr>
<tr>
<td>40-49</td>
<td>6 (32)</td>
</tr>
<tr>
<td>50-59</td>
<td>7 (37)</td>
</tr>
<tr>
<td>60-69</td>
<td>2 (11)</td>
</tr>
<tr>
<td>Working status, n (%)</td>
<td></td>
</tr>
<tr>
<td>Full time</td>
<td>15 (79)</td>
</tr>
<tr>
<td>Part time</td>
<td>4 (21)</td>
</tr>
<tr>
<td>Not working</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Years of mental health work experience, mean (SD; range)</td>
<td>18 (13.7; 1-43)</td>
</tr>
<tr>
<td>Hours of customer work per week, mean (SD; range)</td>
<td>18 (5.9; 9-30)</td>
</tr>
<tr>
<td>Education (multiple options may be chosen), n (%)</td>
<td></td>
</tr>
<tr>
<td>Practical nurse</td>
<td>2 (11)</td>
</tr>
<tr>
<td>Nurse</td>
<td>7 (37)</td>
</tr>
<tr>
<td>Psychologist</td>
<td>11 (58)</td>
</tr>
<tr>
<td>Psychotherapist</td>
<td>7 (37)</td>
</tr>
<tr>
<td>Social worker</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Other</td>
<td>6 (32)</td>
</tr>
<tr>
<td>Context of client work, n (%)</td>
<td></td>
</tr>
<tr>
<td>Specialized health care</td>
<td>12 (63)</td>
</tr>
<tr>
<td>Student health care</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Independent practice</td>
<td>6 (32)</td>
</tr>
<tr>
<td>Clients with..., n (%)</td>
<td></td>
</tr>
<tr>
<td>Mild mental disorders</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Moderate mental disorders</td>
<td>7 (37)</td>
</tr>
<tr>
<td>Severe mental disorders</td>
<td>11 (58)</td>
</tr>
</tbody>
</table>
Table 2. Individual characteristics of the interviewees (n=19).

<table>
<thead>
<tr>
<th>Number</th>
<th>Clinical education</th>
<th>Working context</th>
<th>Role</th>
<th>Clients</th>
<th>Digital tools used and discussed in the interview</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Occupational therapist</td>
<td>Specialized health care</td>
<td>Psychosocial treatment</td>
<td>Adults with mood disorders</td>
<td>TT&lt;sup&gt;a&lt;/sup&gt; and MA&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>2</td>
<td>Psychologist psychotherapist</td>
<td>Independent practice</td>
<td>Psychotherapy</td>
<td>Adults with mood disorders</td>
<td>TM&lt;sup&gt;c&lt;/sup&gt; and TT</td>
</tr>
<tr>
<td>3</td>
<td>Psychologist psychotherapist</td>
<td>Specialized health care</td>
<td>Psychological evaluation</td>
<td>Adults with neuropsychological challenges</td>
<td>TM, TT, WM&lt;sup&gt;d&lt;/sup&gt;, and digital psychological tests</td>
</tr>
<tr>
<td>4</td>
<td>Nurse</td>
<td>Specialized health care</td>
<td>Psychosocial treatment</td>
<td>Adults with psychosis or prodromal symptoms</td>
<td>TM, TT, WM, and digital cognitive rehabilitation therapy</td>
</tr>
<tr>
<td>5</td>
<td>Nurse psychotherapist</td>
<td>Specialized health care</td>
<td>Evaluation and consultation</td>
<td>Youth with psychological symptoms</td>
<td>TM, TT, WM, and MA</td>
</tr>
<tr>
<td>6</td>
<td>Psychologist psychotherapist</td>
<td>Student health care</td>
<td>Psychological treatment</td>
<td>Students with sexuality-related challenges</td>
<td>TT, WM, and digitized questionnaires</td>
</tr>
<tr>
<td>7</td>
<td>Nurse</td>
<td>Specialized health care</td>
<td>Evaluation and psychosocial treatment</td>
<td>Adults with mood disorders</td>
<td>TM, TT, and WM</td>
</tr>
<tr>
<td>8</td>
<td>Psychologist psychotherapist</td>
<td>Independent practice</td>
<td>Psychotherapy</td>
<td>Adults with mood disorders</td>
<td>TT and client-introduced apps</td>
</tr>
<tr>
<td>9</td>
<td>Nurse</td>
<td>Specialized health care</td>
<td>Evaluation and psychosocial treatment</td>
<td>Adults with neuropsychological challenges</td>
<td>TM and TT</td>
</tr>
<tr>
<td>10</td>
<td>Psychologist</td>
<td>Specialized health care</td>
<td>Psychological evaluation and psychosocial treatment</td>
<td>Older adults with psychiatric challenges</td>
<td>TM, TT, and WM</td>
</tr>
<tr>
<td>11</td>
<td>Nurse</td>
<td>Specialized health care</td>
<td>Care coordination and psychosocial treatment</td>
<td>Older adults with psychiatric challenges</td>
<td>TM, TT, WM, and digitized questionnaires</td>
</tr>
<tr>
<td>12</td>
<td>Nurse</td>
<td>Psychiatric inpatient ward</td>
<td>Care coordination and psychosocial treatment</td>
<td>Adults with psychiatric disorders</td>
<td>TT and WM</td>
</tr>
<tr>
<td>13</td>
<td>Nurse</td>
<td>Specialized health care</td>
<td>Evaluation and psychosocial rehabilitation</td>
<td>Adults with psychotic disorders</td>
<td>TM, TT, and WM</td>
</tr>
<tr>
<td>14</td>
<td>Psychologist</td>
<td>Specialized health care</td>
<td>Psychosocial treatment and psychological evaluation</td>
<td>Adults with lowered ability to work</td>
<td>TM and WM</td>
</tr>
<tr>
<td>15</td>
<td>Psychologist</td>
<td>Specialized health care</td>
<td>Psychosocial treatment</td>
<td>Adults with mood disorders</td>
<td>TM, TT, WM, and acted as a DMHI&lt;sup&gt;e&lt;/sup&gt; therapist</td>
</tr>
<tr>
<td>16</td>
<td>Psychologist psychotherapist</td>
<td>Independent practice</td>
<td>Psychotherapy</td>
<td>Adults with psychiatric disorders</td>
<td>TM and TT</td>
</tr>
<tr>
<td>17</td>
<td>Psychologist psychotherapist</td>
<td>Independent practice</td>
<td>Psychotherapy</td>
<td>Adults with psychiatric disorders</td>
<td>TT and client-introduced apps</td>
</tr>
<tr>
<td>18</td>
<td>Psychologist</td>
<td>Independent practice</td>
<td>Neuropsychological rehabilitation</td>
<td>People with neuropsychological problems</td>
<td>TT, WM, and rehabilitation software</td>
</tr>
<tr>
<td>19</td>
<td>Psychologist psychotherapist</td>
<td>Psychotherapy center</td>
<td>Psychotherapy</td>
<td>People with psychological trauma-related problems</td>
<td>TT and therapy centers’ digital materials</td>
</tr>
</tbody>
</table>

<sup>a</sup>TT: teletherapy.
<sup>b</sup>MA: mobile app.
<sup>c</sup>TM: telephone or messaging.
<sup>d</sup>WM: web-based materials.
<sup>e</sup>DMHI: digital mental health intervention.

**Semistructured Interview**

The interview was semistructured [72], focusing on the MHPs’ subjective experiences with digital tools in their professional context in client interaction. Although keeping with the RQs, attention was paid to ensuring that the interviews retained their flexibility and accommodated the variance in the interview contexts. Using the typology by McIntosh and Morse [73], the interview was primarily descriptive and interpretive, focusing on discovering the interviewees’ experiential world as opposed to testing a particular theory aligned with consequent inductive
data analysis. The interview guide is presented in Multimedia Appendix 1.

The first author conducted the interviews remotely using Zoom (Zoom Video Communications). He is a clinical psychologist and service designer experienced in conducting interviews and versed in clinical mental health care. The interviews were recorded after verbally confirming the interviewee’s consent according to national research guidelines) and transcribed verbatim for analysis. The interview durations ranged from 47 to 83 minutes, with an average duration of 56 (SD 9) minutes. The total interview data duration was 18 hours 9 minutes, which led to transcribed materials of 96,707 words.

Inductive Data Analysis

The data analysis was conducted in 2 parts reflecting the hierarchical nature of the RQs: the use of specific digital tools (RQ 1) is subordinate to digital tool use in general (RQ 2). Through this approach, we pursued transparency to the often nebulous theme generation and to establish rigor in the research [74] by showing the relationship between specific digital tool use and higher-order themes. In this study, we defined themes as patterns in the data and followed the definition by DeSantis and Ugarriza [75]: “A theme is an abstract entity that brings meaning and identity to a recurrent experience and its variant manifestations. As such, a theme captures and unifies the nature or basis of the experience into a meaningful whole.” Thus, the theme exhibited both unity across participants and internal variance.

The interview data were analyzed inductively to establish the themes bottom-up instead of deductively testing a particular theory. This aimed to ensure that the interviewees’ perspectives came across in the analysis rather than those of the researchers. We are aware that some components of this analysis approach—such as assessing saturation or following COREQ (Consolidated Criteria for Reporting Qualitative Research) guidelines (refer to later sections)—are not promoted in thematic analysis. Otherwise, the data analysis closely followed the 6-step process by Braun and Clarke [76] described as thematic analysis.

In the first step, “Familiarizing yourself with the data,” the first author transcribed the data verbatim and then confirmed the transcription accuracy by relistenning to the interview tapes with the written transcription, which further familiarized him with the data. In the second step, “Generating the initial codes,” the first author coded all the data using ATLAS.ti software (version 22; ATLAS.ti GmbH), allowing for the initial organization of the data into categories. Then, the first author conducted the third, fourth, fifth, and sixth steps of the analysis—“Searching for themes,” “Reviewing themes,” “Defining and naming themes,” and “Producing the report”—per the 2 RQs. Reflexivity was ensured by the ongoing reflection of code and theme generation through the first author’s clinical background and position and the full context of all the data.

To answer RQ 1, the first author started to search for meanings by categorizing all instances where interviewees discussed specific digital tool use, adding up to 349 codes. For example, all the instances in which the interviewees reflected on the different ways in which they used telephone, SMS text messages, WhatsApp, Skype, and Zoom to stay in touch with their clients were categorized per medium. These categories were further grouped into a higher-order category of “Communication.” In total, 2 other categories were established: tool use related to psychiatric evaluation and diagnostics and tool use to facilitate therapeutic change. Because of the descriptive, pragmatic nature of RQ 1, we chose to report these 3 categories as domain summaries—“summaries of the range of meaning in the data related to a particular topic or ‘domain’ of discussion” [77]. The domains comprised the 3 functions the digital tools served, which were identified from the data. The 3 categories were reviewed to ensure that they included all the digital tools discussed in the interview, and they were named and reported in the Results section.

After analyzing digital tool use for RQ 1, the first author began the development of themes for RQ 2. A total of 335 initial codes included interviewees’ reflections on how they viewed and used digital tools in their client practice in general, and these initial codes were searched for themes. For instance, the recurring notion that digital tools do not replace face-to-face connections was reflected on, similarly to mentions of how digital tools may alleviate resource problems in psychiatry and how it is essential to consider the client’s needs. This search led to the establishment of 3 themes. MHP flexibility in client interaction recurred in almost all interviews, and it was also the most frequent coding category. Thus, the first theme described the client-centered clinical approach that unified the participants. The second theme contrasted with the first by highlighting the variance in the MHP digital toolbox. Finally, a third theme was established by examining how digital tools influenced MHP work. The names of the themes were refined several times to ensure that they captured the essence of the interviewees’ accounts. Particular attention was paid to ensuring that the themes had internal consistency and described the whole data set.

Further efforts were made to ensure that the data analysis was reliable. The first and second authors met 2 times to reflect on the data analysis, review the themes, and name them. The first author translated the interviewees’ quotes from Finnish into English, and another researcher (Maria Vesterinen) reviewed the translations, which led to minor clarifications. The results were annotated with interview and paragraph references (eg, #1:100) to facilitate transparency. Member checking [78] was used to ensure that the interpretations made in the study represented the notions of the MHPs. The draft version of the manuscript was sent to 5 MHPs in October 2022 and November 2022: a total of 2 (40%) MHPs who were interviewed, 2 (40%) MHPs who were employed at HUS, and 1 (20%) independent MHP. Their feedback supported the findings, and only minor clarifications were made based on it. Finally, the researchers confirmed that describing and analyzing the study results conformed to the COREQ guidelines [79].
Results

How Do MHPs Use Different Digital Tools in Client Practice? (RQ 1)

Overview

We found that MHPs used digital tools in client practice for three functions: (1) diagnosis and evaluation and (2) counseling, both of which necessitate (3) communication with the client (Table 3). This evaluation aimed to create an understanding of the client’s challenges and disorders to guide treatment. It typically consisted of interviews and questionnaires complemented with psychological testing when a more thorough understanding of the client’s problems and cognition was required. Counseling sought to alleviate the clients’ symptoms and helped them cope with their challenges. Depending on the MHP’s education and role, it may be psychosocial support in a clinic or psychiatric ward, neuropsychological rehabilitation, short-term therapy, or psychotherapy. In this paper, the term counseling is used to refer to all modes of psychosocial support and treatment. Both evaluation and counseling require contact with the client—communication.

Table 3. The 3 distinct functions that characterize mental health professional digital tool use in client interaction. The functions can be served through analog, digitized, or digital solutions, of which examples are provided.

<table>
<thead>
<tr>
<th>Function</th>
<th>Medium</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communication</td>
<td>Face-to-face interaction and written letters</td>
</tr>
<tr>
<td></td>
<td>Telephone, emails, messaging, and teletherapy with audio and video connection</td>
</tr>
<tr>
<td>Diagnosis and evaluation</td>
<td>Pen-and-paper questionnaires and psychological tests</td>
</tr>
<tr>
<td></td>
<td>Sending questionnaires via email and filling questionnaires on the web</td>
</tr>
<tr>
<td>Creating therapeutic change</td>
<td>Brochures, printed materials, and handouts</td>
</tr>
<tr>
<td></td>
<td>Sharing materials via email or through web-based information portals</td>
</tr>
<tr>
<td></td>
<td>Telettherapy with advanced features such as virtual reality and avatars</td>
</tr>
<tr>
<td></td>
<td>Responsive and gamified tests and integration of various data sources</td>
</tr>
<tr>
<td></td>
<td>Interactive DMHIs(^a), mobile apps, and serious games</td>
</tr>
</tbody>
</table>

\(^a\)DMHI: digital mental health intervention.

We argue that MHP functions reflect the nature of their profession and its practices. Thus, the need for communication with the client, creating an understanding of their challenges through evaluation, and supporting them is likely to remain constant over time, whereas how MHPs achieve these functions may evolve and change. The change is driven by technological advancement, which we describe on a continuum from analog to digitized and digital media. By analog, we refer to nondigital media; by digitized media, we refer to an analog medium converted into digital without substantial changes or additions. In contrast, digital refers to media that use the possibilities native to digital. We acknowledge that the lines between the 3 are not always fully clear and keep evolving; however, the more detailed ontological discussion must be left elsewhere.

The 3 MHP functions may be implemented in analog, digitized, or digital media. Regarding communication, the telephone digitizes verbal interaction, and emails and SMS text messages digitize written communication. In contrast, digital teletherapy solutions can change the nature of the interaction by, for instance, augmenting the conversation with interactive materials and features or placing the meeting in a fictional virtual reality environment with avatars. Concerning evaluation, a pen-and-paper questionnaire can be digitized into a web questionnaire that calculates the results. A digital evaluation solution could enrich these data with psychophysiological measurements, mobile data, and electronic health records, or its execution could be responsive or gamified. The therapy-complementing materials, such as patient guides and CBT worksheets, can be shared via email or on the web. Meanwhile, their digital implementation could, for instance, make use of adaptive elements or take the form of a serious game to make them more engaging and effective. To recapitulate, we assert that the underlying function of the tools remains unchanged even when they grow more interactive, networked, and complex.

The results per function are described as a domain summary. They are as follows: (1) MHPs use complementary channels in client communication, (2) the evaluation of clients is being digitized, and (3) MHPs support therapeutic change using digital materials.

MHPs Use Complementary Channels in Client Communication

Face-to-Face Interaction

MHPs found many unique benefits with face-to-face interaction. Unmediated contact allowed for a superior connection with the client as the MHP could observe and react to nuances in client expressions and behavior that would otherwise be lost. This also enabled the MHP to generate a more reliable and accurate understanding of their problems, which was also valuable for psychological evaluation. Coming to the meeting in person also activated the client, which was found to be beneficial for clients with a tendency toward isolation and passivity. Finally, the in-person social interaction can be therapeutic in itself. An MHP facilitating group therapy for clients with social anxiety described the following:

*The significant exposure is that you come to the group in person, and that you spend time with other people.*

[\#15:62]

The prevailing sentiment regarding the value of face-to-face meetings was succinctly described by an MHP working in psychiatry:

https://humanfactors.jmir.org/2023/1/e44681
Of course, it [teletherapy] will never replace it [face-to-meetings]. We see a lot more than a person’s face when they arrive [to the practice]; there is the presence, the whole person.[#7:115]

**Telephone and SMS Text Messaging**

MHPs used the telephone and SMS text messages to schedule meetings and checkups on their clients, and the telephone was also occasionally used for counseling. Only 2 MHPs reflected on the therapeutic potential of asynchronous messaging. One of them provided low-threshold support to their clients via WhatsApp, finding that merely exchanging messages could help them through a challenging situation and alleviate anxiety. An MHP with an occupational focus found that the time-independent nature of messages enhanced in-person therapy and allowed the therapy to “live in the mind” of their client between sessions. However, most MHPs’ client interactions occurred in scheduled meetings.

**Teletherapy**

The COVID-19 pandemic led many MHPs to convert some in-person meetings to a remote format, a practice that prevailed even after the pandemic. “The remote therapy has come to stay” (#15:39), as summarized by an MHP. Few MHPs explicitly preferred face-to-face meetings and were reluctant to schedule remote ones. The MHPs found remote meetings flexible and that they had the benefit of saving the client travel time. Remote meetings also facilitated a larger number of participants, both clients and MHPs, also from different locations. The decision between face-to-face therapy and teletherapy was often influenced by the client’s preference rather than readiness factors such as having a computer with a video camera and competence to use them. The readiness factors were only emphasized with some geriatric clients.

Interestingly, holding the meetings remotely did not necessarily change their content, indicating that they were digitized communication rather than natively digital. The videoconference meetings were found to be mediated counseling where, for instance, screen sharing had a similar function to a whiteboard in an in-person meeting. When the clients’ problems were not considerably debilitating, they could reflect on their behavior and be present in the relationship; remote counseling occurred very similarly to face-to-face meetings. An MHP conducting long-term psychotherapies reflected the following:

> I can report that the therapy meeting works pretty much the same way. When I think back on the sessions, I don’t perceive a difference whether the session was conducted remotely or in person because the very same things happen, and it works in the same way. #[17:38]

To summarize, we found that MHPs used media—face-to-face meetings, phone calls, messaging, and teletherapy—complementarily to serve different needs in the therapeutic relationship. The client preference and readiness influenced the medium chosen. However, it appeared that the different tools typically digitized the established practices rather than changing their content substantially.

**The Evaluation of Clients Is Being Digitized**

**Overview**

MHP responsibilities often included evaluation and counseling (Table 2). Psychiatric care routinely began with an evaluation period and continued with counseling. However, the balance between the 2 varied—some MHPs offered mainly counseling services, and one concentrated solely on psychological evaluation. Others blended evaluation and counseling in their work.

**Pen-and-Paper Materials**

MHPs routinely used pen-and-paper symptom questionnaires such as the Beck Depression Inventory [80] and Clinical Outcomes in Routine Evaluation-Outcome Measure [81] to evaluate the clients’ symptoms and track treatment progress. When a more extensive evaluation was needed, particularly on the client’s cognition, psychological tests were conducted, and in the sample, they were performed exclusively on pen and paper. Some organizations had “digitized” the questionnaires impromptu because of the coronavirus pandemic—they were sent to the clients via SMS text message or email, a practice that one MHP considered questionable.

**Digital Platforms**

A more sustainable solution for managing questionnaire data came from digital platforms that facilitate the collection, management, and storage of client data. Many organizations were picking up new systems to facilitate their work; however, the progress in their implementation varied alongside the MHP experiences of them. Some found the platforms beneficial and useful, whereas others were not equally impressed by their unwieldy implementation or were concerned that their older clients could not use them without support.

The study found that the digitalization of questionnaires and client data is underway in many organizations. Compared with the adoption of communication software, which was found to be necessary for the work, the adoption of questionnaire software appeared slower. In addition, the analog pen-and-paper materials that the MHPs were accustomed to using did not propose considerable drivers for change.

**MHPs Support Therapeutic Change With Digital Materials**

**Overview**

We found that most MHPs used some type of material to augment the effectiveness of counseling. The materials were used to give the client information—psychoeducation—on their disorder or condition, such as depression, anxiety, psychosis, sleeping, or pain. This aimed to develop the client’s confidence and capability to self-manage the symptoms. “We have to help the client to help themselves” (#14:127), explained an MHP on their philosophy regarding the materials. The visual materials also complemented the discussion-oriented therapeutic contact. They gave the client “something else than just talk” (#11:154): concrete materials and tools to use.

MHPs found that their clients’ reactions to the materials varied considerably. An MHP summarized the following:
Some think they are completely rubbish, useless slips of paper. Others think they are lovely: they appreciate that there is something concrete. [#11:153]

An explanation for this difference was the variance in the clients’ interest in reading and their capability for self-reflection. To facilitate the adoption of the materials, the MHPs often presented them in the session, encouraged the client to explore them after the meeting at their own pace, and followed up on them.

**Analog Materials**

Some MHPs handed out paper brochures on disorders, and others handpicked materials from the internet or their own resources. The rationale for printing out or photocopying the materials was to make them more tangible and understandable, help the clients who do not have competency in finding the materials on the web themselves, and encourage clients to read the materials.

**Digitized Materials**

Almost all MHPs recommended digitized materials to their clients at least occasionally. The most common resource was the national MentalHub, which includes materials per disorder with separate content for youth. Its modular structure was found to be convenient, and the interactive sections and questionnaires were appreciated. In addition, third-sector services and materials, videos, and handpicked materials were recommended.

**Mobile Apps**

The MHPs rarely used or recommended mobile apps to their clients. When apps were used, the MHPs were personally familiar with them or the app was published by a credible public organization and targeted to the MHP clientele, as was the case with the youth-targeted Chillaa app.

**DMHI Materials**

Many MHPs had an indirect experience with DMHIs, most commonly with the national DMHI prescribed to their clients. They found that it could facilitate access to therapy for clients in sparsely populated areas, complemented the MHP know-how in specific domains, and was an option for new clients as they waited for therapist contact for “2-3 months” (#13:125) or “4 months” (#6:105).

Despite the advantages, the MHPs were generally at least somewhat hesitant and cautious regarding DMHIs. The nontransparent nature of the contents of the national DMHI discouraged it from being recommended. The MHPs perceived that DMHIs were most suitable for clients with relatively mild psychiatric problems, such as subclinical anxiety, stress, or relationship challenges. In contrast, many MHPs worked in special health care or with clients who had considerable clinical challenges and sought help because they could not manage their behavior without support. Thus, the clients needed and expected face-to-face reflection—“the ears and voice of the other person” (#6:119). The MHPs found that interpersonal contact—the therapeutic alliance between the MHP and the client—was constitutional, vital, healing, and remedial, and the lack of human interaction was the primary concern of almost all MHPs regarding DMHIs. If DMHIs were used, MHPs explicitly preferred supported over unsupported interventions. An MHP conducting long-term psychotherapy described the following:

*Personal contact is of utmost importance. I believe that everyone may not need it, but the majority do. Some may get, at least for some time, help and relief from their socialization with a machine but it cannot replace a human.* [#2:145]

MHPs actively used materials in their client interactions. Across the media, the function of the materials was to help the client gain insights into their symptoms and ways to manage them. Therefore, we propose that the therapy-supporting materials could be viewed in a continuum from analog brochures to digitized self-help materials to (therapist-supported) structured DMHIs.

**What Characterizes the MHPs’ Digital Tool Use in Client Practice in General? (RQ 2)**

**Overview**

After analyzing how MHPs used particular digital tools, we examined their digital tool use in general in the context of their client work. The analysis of the interview data established three themes: (1) digital tool use is negotiated in client interaction, (2) autonomy and contexts diversify MHPs’ digital toolbox, and (3) existing practices invite incremental developments (Figure 1).

The three themes correspond with the nature of MHPs’ work, which we characterize as (1) client-centered, (2) independent, and (3) a service. MHPs exhibited client-centricity by being closely mindful of their clients and adjusting their digital tool use according to the perceived needs of their clients. This negotiation was shaped by the possibilities in the MHP toolbox. As MHPs had independence and autonomy in compiling their toolboxes, there was considerable heterogeneity in their contents and breadth. Third, the MHP work was a service—it focused on intangible interaction, was difficult to standardize, was produced and consumed simultaneously, and could not be stored [82]. The MHPs perceived that the digital tools augmented the interpersonal service they offered and, therefore, offered incremental rather than radical developments to the existing practices.

We describe how the MHP work was influenced by 3 layers of context. The clinical context included practices and expectations for flexible interpersonal MHP-client interaction. The organizational context may provide the MHP with tools such as digital platforms and processes for using them. The broader technological and cultural developments influenced the digital tools available in society and the MHP and client willingness and competency in using them.
Digital Tool Use Is Negotiated in Client Interaction

Almost all MHPs highlighted how they adjusted their behavior to their clients’ individual situations, needs, and symptoms in both evaluation and counseling. They adjusted, for instance, the focus of the evaluation; its duration; the frequency of counseling; and the therapeutic techniques, exercises, questionnaires, and digital tools used. MHPs may, for instance, offer their clients the possibility of choosing between face-to-face and remote meetings and whether a particular therapy was conducted in pen and paper or digitally assisted. Rather than following a rigid care routine, the MHPs found it vital to tailor the interaction to the client in the moment.

MHPs found negotiation and flexibility necessary because of the considerable variance in their clientele. Although they may work with a particular group of people, there was still substantial variation in the life context, symptoms, and needs of their clients; moreover, the clients’ situations may fluctuate. Thus, the key question in counseling was “finding the right tool at the right time” (#17:95). This position also reflected the nature of psychological problems—the MHP cannot directly influence the behavior of the client, who is ultimately responsible for the change. Flexibility also meant fostering client autonomy. An MHP described the following:

I have the overall approach that I offer the client different means, tools, and then they decide. [#1:83]

Supporting client autonomy was also exhibited in how using a digital tool can be client-initiated and how MHPs considered client readiness and preferences in their tool recommendations. Overall, digital tools were perceived to serve higher-order therapeutic aims; they were “a means to an end” (#5:202).

Our research showed how MHPs prioritized establishing a working therapeutic relationship with their clients by adapting their behavior and the tools used. This reflected a client-centered profession, position, and practice.

Autonomy and Contexts Diversify the MHP Digital Toolbox

The MHP can only suggest exercises, materials, and tools that they know of, can access, and perceive as beneficial. The digital possibilities at the MHP disposal are referred to in this paper as the MHP digital toolbox, in which we found considerable variance. On one end, the digital toolbox was considerably limited—the MHPs used digitized and digital tools only to communicate with the client. “The only time when electricity flows through the wires is when we use the telephone” (#14:195), expressed an MHP of their nondigital care pathways. On the other end, MHPs used a breadth of communication channels, digitized and digital materials, and even apps and were aware of the national DMHI.

MHPs have the autonomy to compile their toolboxes. “When I am travelling, I may pick up something that I find works well for rehabilitation purposes: it can be a booklet, a game, or whatever” (#18:201), reflects an MHP. Thus, the MHPs’ attitude toward, interest in, and experience with digital solutions influenced the breadth of their digital toolbox. Those with more experience with digital tools showed higher competence and more positive attitudes toward them. Some MHPs proactively reflected on how they were generally curious about new digital tools, sought training on them, explored them on their own, and even participated in their development. Others exhibited a far more restricted and cautious stance on digital tools and self-perceived their digital skills as low and avoided their use.

Although MHPs have the autonomy to shape their digital toolbox, their freedom is limited and influenced by the possibilities in their organizational and societal context. Larger organizations often provided the MHPs with communication...
and questionnaire platforms and restricted them to the chosen platform, whereas MHPs working in their own practice had more liberty to choose these tools. The availability of credibly perceived digital tools, such as MentalHub psychoeducational materials, national DMHIs, third-sector resources, and some apps, encouraged their uptake. The external societal context also influenced the MHPs through their clientele—some clients requested remote meetings and introduced apps in counseling. Many MHPs found that their clients had the competency and means to use telehealth channels and they could search for and access digital content with little guidance, which was related to the digital tools being broadly used in Finnish society.

Digital tool use was influenced by the breadth of the MHP digital toolbox that the MHP can compile independently. Its contents were limited by the possibilities in the MHP organizational and societal context, and MHP attitudes, preferences, and experiences influenced tool uptake.

**Existing Practices Invite Incremental Developments**

Aside from the national DMHI, we found that the presently used and emerging tools brought incremental developments to the client practice. The teletherapy solutions reduced travel times but did not change the nature of counseling itself, the digitization of the pen-and-paper questionnaires allowed the same instruments to be filled on the web, and digitized psychoeducational materials served the same purpose as analog handouts. In other words, the digitized tools allowed the MHP to perform the tasks they already performed in an analogous manner but more effectively. These developments retained the nature of MHP work as a service; did not offer the scalability benefits expected from digital interventions; and, therefore, did not directly address the insufficient resources in mental health care, which many MHPs were conscious of.

The MHPs’ principal hesitancy regarding digital treatments was their perceived insufficiency for their clients. Most MHPs highlighted how their clients had severe challenges and needed therapist interaction and that unsupported digital treatments were best suited for those with mild challenges. In addition, their clients sought counseling and expected interpersonal contact rather than a digital solution. Some MHPs were aware of how their position in the mental health ecosystem may have affected their thinking and attitudes. “This may be associated with my position in the treatment and service chain. It brings the view that [DMHI] was not enough and that what is needed is something longer and more intensive” (#8:144), an MHP pondered. In general, MHPs viewed that their professional service could not be replaced by a digital tool as its core lay specifically in human interaction.

Despite limitations, many MHPs could imagine the benefits that digital tools can offer in the future. They could extend the reach and access to therapy services in remote areas and offer specialized therapy services. They could lower the threshold to seek help and be helpful to clients who withdraw from others, are anxious and uncommunicative in MHP interaction, or have difficulties reflecting on their emotions verbally. Digital content could be more attractive; engaging; experience-rich; interactive; flexible; and, therefore, more effective than analog materials. Through their presence in the client’s everyday life, digital interventions may provide flexible support whenever and wherever the client wants and needs it; they could activate the client through notifications and adaptive exercises and give them encouraging, timely feedback. Overall, the digital possibilities could expand the treatment portfolio:

*We need a large toolbox if we want to help everyone because people are so different. Also, they have different needs, different skills, and different capabilities to participate in something. So, we need a toolbox with a wrench, a screwdriver, all sorts of things.* [#16:221]

In summary, the digital tools implemented in existing client interaction—emphasizing practices were likely to contribute to incremental developments in MHP work. Although many MHPs were interested in new digital tools, their perspective was limited by their clientele, client expectations, and the nature of their work as a service.

**Discussion**

**Principal Findings**

It has been suggested that digital tools may aid in closing the mental health treatment gap. Unfortunately, even if effective, many new interventions fail in their implementation in complex real-life environments [83]. To facilitate this change, it is necessary to understand how MHPs use digital tools in clinical practice. Our study showed that digitized and digital tools were becoming a part of the clinical practice—MHPs used various channels in client interaction, the evaluation and diagnosis of clients were being digitized, and web-based materials were frequently used to supplement counseling. The MHPs used the tools flexibly, adapting to their clients; there was variance in the breadth of MHP digital toolboxes; and the tools offered primarily incremental developments in MHP practice. These findings have vital implications for developing and implementing digital tools in mental health care.

**Contributions to Existing Research**

Our research exhibited how MHPs use multiple channels to communicate with their clients. Most interviewed MHPs engaged their clients face-to-face and through telecounseling, a change that the COVID-19 pandemic has expedited [20-22]. Previous research has found that teletherapy offers convenience and flexibility [20], which we also identified as a motivator. We enrich previous findings by highlighting how the platforms for communication may be flexibly chosen per client and that there was considerable variance in MHP practices. Regarding messaging, for instance, only 2 MHPs explained how they used asynchronous messaging to support their clients. This finding invites consideration of how MHP services could be systematically augmented with, for instance, messaging [84] to improve treatment adherence and monitoring and offer support.

Previous research has found that MHPs perceive DMHIs as complementary to face-to-face therapies, prefer blended over stand-alone digital interventions, and consider unguided interventions insufficient for clients with substantial challenges [8]. We had similar findings and suggest that they may be explained by MHPs viewing the core of their work as lying in...
client interaction, where technologies hold a secondary, supportive role. This position reflects the history and nature of the profession and is supported by studies that have found the therapeutic alliance to be an essential common factor for therapeutic outcomes [12,85,86]. This also suggests that the digital transformation in mental health care may be driven by client and systemic needs rather than by MHP-driven motivators. The former includes the growing demand for mental health care, improving access to treatment in underserved areas, and offering an alternative to interpersonal treatments.

MHP attitudes toward digital tools may be explained by their experiences with them. Indeed, the 2 go hand in hand—attitudes are related to tool use [87,88], and use is related to more positive attitudes toward digital tools [9]. The direction of causality, however, remains unclear: do MHPs who consider the tools more positively use them more often, or do those who begin to use them grow more positive in their attitudes? Regardless, our research complements the findings by describing how positive attitudes and experiences might manifest in a broader digital toolbox whose contents can be used flexibly in client interaction. As the digital transformation progresses, this may lead to a growing divide between MHPs with extensive digital tools at their disposal and those who double down on the face-to-face approach.

The relationship between MHP attitudes and digital tool use may be conceptualized using the theory of planned behavior [89]. It posits that positive intentions are associated with the likelihood of the associated behavior occurring. Positive intentions, in turn, are influenced by the favorability of attitudes, positive social norms, and perceived ease or difficulty in performing the behavior. The theory has found empirical support in studies of MHP attitudes [87,90,91]—professionals are more likely to use digital tools when they view them favorably, their peers use them, and they find the tools easy to use, and these 3 factors can offer conceptual avenues for facilitating tool implementation.

Implications for Digital Tool Development

The dominant mental health service delivery model reflects its psychotherapeutic roots—highly trained professionals offer services in one-on-one in-person settings [92]. However, as the need for mental health services grows, there is ever more awareness of the limitations of the delivery model—it lacks scalability as it is closely tied to a scarce human resource, MHP time. However, uncoupling time from the service delivery challenges the very fundament the services are based on, the interpersonal therapeutic alliance [11]. Our research suggests that MHPs may view telehealth solutions and digitized questionnaire suites more favorably than digital interventions as the former complements personal interaction, whereas the latter challenges it. We surmise that these attitudes may extend to other technological developments such as artificial intelligence, chatbots, social media, and virtual reality [7,93], which may be viewed as disrupting the beneficial qualities of the therapeutic relationship unnecessarily.

We suggest consciously distinguishing 3 modes of treatment (Figure 2) to alleviate the tension between them. Psychosocial interventions such as counseling are founded on interpersonal interaction and the therapeutic relationship [11,85,86]; biomedical treatments such as psychiatric medication and electroconvulsive therapy affect the nervous system; and independently used interventions such as psychoeducational courses [23] and digital interventions [94] are based on clients acquiring new skills, building motivation, and creating opportunities to change [95]. Psychosocial interventions are flexible, interpersonal, and adaptive to the client, but on a societal scale, their benefits are tied to the available MHP resources. Psychiatric medications are scalable but may have side effects. Independently used interventions can offer best practice psychoeducation and guidance at a scale whenever and wherever, yet they lack the empathetic and motivating interpersonal connection.

Differentiating between the different modes of treatment can help clients and clinicians perceive their complementary potential. Associating an independently used intervention with a web-based program or course instead of therapy may help establish realistic expectations and facilitate reaching clients who expect and benefit from such an approach. Interestingly, we found that many MHPs already encouraged their clients to access analog or digitized materials between sessions. Thus, some MHPs appear to have adopted a blended approach [36] where the treatment uses both psychosocial and publicly available, independently used components. However, the self-adopted practices lacked consistency, which invites the consideration of systemic ways to improve service processes.

Focusing on psychosocial interventions, our model (Table 3) allows for the differentiation of which function—communication, evaluation and diagnosis, and facilitating therapeutic change—the digital tool relates to. This allows developers and health care management to advance a user-centered position and connect the solution with existing practices and MHP and client needs. However, further research is needed to understand how to harmonize the 3 modes of treatment in clinical practice so that they complement each other throughout the client journey.

Training and education are commonly recommended to improve MHP digital tool adoption and use [10,39,96-99]. This individual-focused approach is synergistic with the autonomy and independence that MHPs enjoy in their work, which may also diminish the impact of group- or organizational-level change efforts [100]. We maintain that training may be sufficient to implement digitized tools that suggest incremental changes in clinical practices. In contrast, the digital tools that present radical changes to the modus operandi must be accompanied by substantial structural changes. Several frameworks may be helpful in this regard [101]. They include the Consolidated Framework for Implementation Research [8,83]; Promoting Action on Research Implementation in Health Services [102,103]; and the Nonadoption, Abandonment, Scale-up, Spread, and Sustainability framework [104,105]. All these frameworks highlight how creating a usable digital tool is not enough—it needs to be considered in terms of the adopters, clinical practices, organizational care processes, and societal context. Our work complements these models by illuminating the MHP-client implementation context (Figure 1).
Figure 2. Differentiating between 3 modes of treatment. Psychosocial interventions are based on interpersonal interaction, biomedical interventions affect the nervous system, and independently used interventions encourage change through learning.

Limitations
This research was conducted in Finland, and the results showed that the external context influences MHP behavior and attitudes. Finland is highly developed digitally [106]. Of the adult population, 93% use the internet [107], and web-based interaction with governmental services is very common [106]. In health care, e-services, including e-prescriptions, are commonly used, and public and many private services have integrated their patient data into a shared repository called Kanta [108]. The status of DMHIs for depression has been legitimized by their acceptance into the national clinical practice guidelines [109]. Aligned with the theory of planned behavior [89], the societal digital development and broad use of digital devices in health care are likely to influence MHP and client attitudes positively and contribute to their adoption.

The study recruitment efforts sought to attract participants from various backgrounds as well as those who held a more critical stance on digital tools. However, the sample included only a few critical voices and several MHPs with substantial knowledge of digital tools and therapies. The research theme likely attracted those with a more positive stance on the topic and may not represent the entire MHP population. The interviewees included psychologists, psychotherapists, nurses, and an occupational therapist who worked in various contexts, with clients of different ages, and with clients who had various disorders. One MHP was experienced in conducting a national DMHI. The sample did not include physicians or MHPs from private occupational health care or basic-level health care for nonstudents, providing avenues for future research efforts. In addition, the study does not necessarily reflect the positions of the health care leadership who may not directly work with clients and are responsible for managing and developing the service and information systems used within them. Further research may be needed to understand the leadership position and strategy and consider the change in management efforts required to implement the solutions in organizational settings [109].

Conclusions
New digital tools are being actively developed in mental health care, and scalable solutions are expected to alleviate the global mental health problem. This study illuminated the context of MHPs, who play a crucial role in adopting and implementing new technologies in client interaction. Our research showed that MHP work involves 3 key functions: communicating with the client, diagnosing and evaluating them, and facilitating therapeutic change. Teletherapy was widely accepted and adopted alongside other media, and the evaluation of clients was becoming more digitized. Digitized psychoeducational materials were widely used, but MHPs hesitated to recommend stand-alone digital therapies that were perceived as insufficient for their clients.

We characterized the MHP work as a client-centered independent service. The MHPs adjusted the techniques, interactions, and tools used per client. The MHPs had the independence and autonomy to choose the tools used from a range of possibilities. This created heterogeneity in the digital toolboxes, which were influenced by MHP preference, organizational context, and tools available in the external environment. The digital tools introduced to this service context often proposed incremental rather than radical developments, considerably limiting their impact. More research is needed to examine when and how scalable, independently used interventions can best complement psychosocial interventions and for whom they may work independently.
Acknowledgments

LL and JMP are funded by the Jane and Aatos Erkko Foundation and Technology Industries of Finland Centennial, The Future Makers Program, and Business Finland Research2Business (42173/31/2020). VMK is funded by the Academy of Finland (grant 353267) and the European Research Council under the European Union Horizon Europe research and innovation program (grant 101042052). The authors would like to thank all the mental health professionals who participated in the research, Saara Tikka and Sini Oksanen for their help in the transcriptions of the first 5 interviews, and Maria Vesterinen for her valuable input with the translation of the quotes.

Authors’ Contributions

LL designed the study, conducted the interviews, transcribed them, conducted the interview data analysis, wrote the initial version of the manuscript, and revised it. VMK contributed to the analysis of the data and revised the manuscript. JMP contributed to the approach and structure of the paper and revised the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The interview guide.

References


46. Clarke J, Draper S. Intermittent mindfulness practice can be beneficial, and daily practice can be harmful. An in depth, mixed methods study of the “calm” app’s (mostly positive) effects. Internet Interv 2019 Nov 16;19:100293 [FREE Full text] [doi: 10.1016/j.invent.2019.100293] [Medline: 31890639]


64. Lukka L, Karhulahti VM, Palva M. Mental health professionals’ views and needs regarding game-based digital therapeutics. Open Science Framework. 2022. URL: https://osf.io/b8eg6 [accessed 2023-04-17]


88. Feijt MA, de Kort YA, Bongers IM, Ijsselsteijn WA. Perceived drivers and barriers to the adoption of eMental health by psychologists: the construction of the levels of adoption of eMental health model. J Med Internet Res 2018 Apr 24;20(4):e153 [FREE Full text] [doi: 10.2196/jmir.9485] [Medline: 29691215]


Adult Patients’ Experiences of Using a Patient Portal With a Focus on Perceived Benefits and Difficulties, and Perceptions on Privacy and Security: Qualitative Descriptive Study

Elisa H Son1*, RN, PhD; Eun-Shim Nahm2*, RN, PhD

1Translational Biobehavioral and Health Disparities Branch, National Institutes of Health Clinical Center, Bethesda, MD, United States
2Organizational Systems and Adult Health, University of Maryland School of Nursing, Baltimore, MD, United States
*all authors contributed equally

Corresponding Author:
Elisa H Son, RN, PhD
Translational Biobehavioral and Health Disparities Branch
National Institutes of Health Clinical Center
10 Center Drive
Bethesda, MD, 20892
United States
Phone: 1 667 701 5768
Email: hyojin.son@nih.gov

Abstract

Background: Patient portals can facilitate patient engagement in care management. Driven by national efforts over the past decade, patient portals are being implemented by hospitals and clinics nationwide. Continuous evaluation of patient portals and reflection of feedback from end users across care settings are needed to make patient portals more user-centered after the implementation.

Objective: The aim of this study was to investigate the lived experience of using a patient portal in adult patients recruited from a variety of care settings, focusing on their perceived benefits and difficulties of using the patient portal, and trust and concerns about privacy and security.

Methods: This qualitative descriptive study was part of a cross-sectional digital survey research to examine the comprehensive experience of using a patient portal in adult patients recruited from 20 care settings from hospitals and clinics of a large integrated health care system in the mid-Atlantic area of the United States. Those who had used a patient portal offered by the health care system in the past 12 months were eligible to participate in the survey. Data collected from 734 patients were subjected to descriptive statistics and content analysis.

Results: The majority of the participants were female and non-Hispanic White with a mean age of 53.1 (SD 15.34) years. Content analysis of 1589 qualitative comments identified 22 themes across 4 topics: beneficial aspects (6 themes) and difficulties (7 themes) in using the patient portal; trust (5 themes) and concerns (4 themes) about privacy and security of the patient portal. Most of the participants perceived the patient portal functions as beneficial for communicating with health care teams and monitoring health status and care activities. At the same time, about a quarter of them shared difficulties they experienced while using those functions, including not getting eMessage responses timely and difficulty finding information in the portal. Protected log-in process and trust in health care providers were the most mentioned reasons for trusting privacy and security of the patient portal. The most mentioned reason for concerns about privacy and security was the risk of data breaches such as hacking attacks and identity theft.

Conclusions: This study provides an empirical understanding of the lived experience of using a patient portal in adult patient users across care settings with a focus on the beneficial aspects and difficulties in using the patient portal, and trust and concerns about privacy and security. Our study findings can serve as a valuable reference for health care institutions and software companies to implement more user-centered, secure, and private patient portals. Future studies may consider targeting other patient portal programs and patients with infrequent or nonuse of patient portals.

JMIR Hum Factors 2023;10:e46044 doi:10.2196/46044

https://humanfactors.jmir.org/2023/1/e46044
Introduction

Background

Patient engagement in care management, such as informed choices and shared decision-making, is emphasized in current health service delivery [1,2]. The widespread use of health information technologies (HITs) has enabled patients to be more actively involved in their care activities [3]. A patient portal is a type of HIT linked to electronic health records (EHRs) that allows patients to view their medical records, communicate with care providers, and perform other care-related tasks [4]. Driven by national efforts over the past decade, patient portals are being widely implemented by hospitals and clinics across the nation [5].

A large body of patient portal research has focused on investigating the effects of using patient portals or factors that may influence patient portal adoption. Researchers have demonstrated the positive effects of using patient portals on patient engagement in care activities (eg, appointment adherence and medication management) and clinical health outcomes (eg, blood pressure and blood glucose control) [6,7]. Regarding factors associated with patient portal adoption, perceived usability has been discussed as a main factor [8-10]. Common usability issues included problems with log-in or access and difficulties in understanding information or navigating functions in patient portals [8,9]. Another notable factor associated with patient portal adoption was concerns about privacy and security risks [11]. Primary concerns were disclosure of personal health information to others outside of one’s permission or unauthorized use of the information by third parties [11]. While usability issues and privacy concerns were the main barriers, health care providers’ recommendations and training support were facilitating factors of patient portal adoption [12,13].

According to a national survey (N=3865) conducted in 2020, about 60% of people were offered digital access to medical records, and about 64% of them accessed their medical records in the past 12 months [14]. As the percentage of patients using patient portals increases, the perception and experience of using patient portals may vary among the users from diverse demographic and clinical backgrounds [8,15]. Continuous evaluation of patient portals and reflection of feedback from end users across care settings are needed to make patient portals more user-centered after the implementation. Existing studies examining patients’ experiences of using patient portals often included small samples recruited from limited clinical or research environments [2,8,16]. There still is a lack of empirical understanding of the lived experience of using patient portals in larger samples recruited across care settings.

Objective

To fill the current gap in patient portal research, we conducted an anonymous digital survey of adult patients recruited from a variety of care settings to examine their comprehensive experiences of using a patient portal [17]. As part of the survey, the participants submitted qualitative comments on their perceived benefits and difficulties of using the patient portal and perceptions on privacy and security. The aim of this study was to investigate the lived experience of using the patient portal in adult patients recruited across care settings, focusing on their perceived benefits and difficulties of using the patient portal, and trust and concerns about privacy and security.

Methods

Study Design and Participants

This qualitative descriptive study was part of a cross-sectional digital survey research to investigate the comprehensive experience of using a patient portal in adult patients who had accessed the portal in the past 12 months [17]. The participants were recruited from 20 care settings in hospitals and clinics of a large integrated health care system in the mid-Atlantic region of the United States; the selected care settings represented various geographical locations (urban and rural), treatment areas (primary and special), and patient portal activation densities (large and low). Inpatients or outpatients who were 18 years or older, had an active patient portal account offered by the health care system (MyChart by Epic Systems Corporation), and had used the portal at least twice in the past 12 months prior to the survey were eligible for participation. A 1-time anonymous survey was administered from August 19 to September 20, 2019. An eMessage that includes a hyperlink to the web-based survey with a brief invitation was sent to 9949 patients who had visited the selected care settings a week prior to the start of the survey. A total of 743 patients participated in the survey, and data from 734 patients who responded to at least one open-ended question were included in the current analysis.

Selected Data and Measures

Demographic and descriptive variables included age, sex, race or ethnicity, marital status, education, monthly income, employment, presence of chronic disease, and internet usage hours per week. eHealth literacy was measured using the eHealth Literacy Scale, an 8-item 5-point Likert scale (1=strongly disagree; 5=strongly agree) [18]. The eHealth Literacy Scale is internally consistent and valid [19,20], and it had a Cronbach α of .93 in this study. The perceived usability of the patient portal was measured using a modified Perceived Health Web Site Usability Questionnaire, a six-item 7-point Likert scale (1=not at all worried; 7=very worried) adopted from the National Consumer Survey on HIT conducted for California HealthCare Foundation [22]. The frequency of patient portal use in the past 12 months was measured using a single ordinal item (1 to 9 times, about monthly, more than monthly).

https://humanfactors.jmir.org/2023/1/e46044

JMIR Hum Factors 2023 | vol. 10 | e46044 | p.681
(page number not for citation purposes)
The participants were further asked open-ended questions about beneficial aspects of the patient portal, specific difficulties that they experienced when using the patient portal, and their trust in privacy and security of the patient portal. For those who answered that their level of concern about privacy and security of the patient portal was high (i.e., a score higher than 4 out of 7), they were additionally asked why they were concerned about privacy and security [22].

Data Analysis

Descriptive analysis was performed using SPSS (version 28; IBM Corp) for each demographic and descriptive variable including mean, SD, frequency, and percentage.

A total of 1589 qualitative comments were collected on 4 open-ended questions: beneficial aspects of the patient portal (734 comments); difficulties in using the patient portal (179 comments); trust in privacy and security of the patient portal (554 comments); and concerns about privacy and security of the patient portal (122 comments). A combination of inductive coding and content analysis was conducted to elicit the main themes from the qualitative comments [23,24]. A set of coding rules were defined prior to the initial coding. The coding unit was a sentence, and the context unit was a question for each topic. The qualitative comments were coded into mutually distinct themes, and the frequency of coding units was calculated for each theme. If multiple sentences in a single comment referred to the same concept, the sentences were coded once as one unit. When a sentence included more than 1 concept, the sentence was coded multiple times for the applicable themes.

Using an Excel (Microsoft Corp) spreadsheet, 2 coders individually performed the initial coding following the same coding rules. The coders were nurse researchers with doctorate degrees who had conducted and published qualitative studies several times. Despite the use of predetermined coding rules to define the context unit and the coding unit, there was no overarching predetermined coding framework. Each coder independently derived mutually exclusive themes that emerged from the qualitative data. The coders compared their thematic results and discussed any coding discrepancies until a consensus was reached. As the coding progressed, the derived themes and unit frequencies were reviewed and refined through iterative discussions between the coders. A total of 22 themes on the 4 topics were finalized: beneficial aspects of the patient portal (6 themes); difficulties in using the patient portal (7 themes); trust in privacy and security of the patient portal (5 themes); and concerns about privacy and security of the patient portal (4 themes). Reliability related to the interpretation of word meanings was ensured by setting clear coding rules and following an iterative approach throughout the analysis. Semantic validity was achieved by assessing the correspondence between the categorization of the coding units and the question topics.

Ethics Approval

Patients invited to the survey were able to decide to participate voluntarily. The survey research was approved by the institutional review board of University of Maryland, Baltimore (HP-00084885).

Results

Demographic and Descriptive Characteristics

Table 1 represents the demographic and descriptive characteristics of our sample (N=734). The mean age of the participants was 53.1 (SD 15.34, range 18-92) years. The majority of them were female (67.6%), White (68.5%), non-Hispanic (97.4%), and had some college or higher degree (83.3%). About two-thirds of them were married or living with a partner (62.9%), employed either full-time or part-time (57.7%), and had a monthly income of US $3000 or higher (60.2%). The majority of them reported having at least 1 chronic disease (86.5%); high blood pressure (46.8%) and high cholesterol (37.4%) were the 2 most reported chronic diseases.

On average, the participants used the internet 24.9 (SD 20.78) hours per week. They showed relatively higher mean scores of eHealth literacy (mean 31.2, SD 5.51) and perceived usability of the patient portal (mean 36.6, SD 6.00) compared to previous studies with older adult web-based users [25,26]. Slightly less than half of them (47.3%) used the patient portal monthly or more frequently during the past 12 months. On a scale of 1 to 7, the mean score for a single item measuring concerns about privacy and security of the patient portal was 2.7 (SD 1.81); 18.7% selected a score value 5, 6, or 7, indicating “worried.”
Table 1. Demographics and descriptive statistics (N=734).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>53.1 (15.34)</td>
</tr>
<tr>
<td>range</td>
<td>18.0-92.0</td>
</tr>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>209 (32.4)</td>
</tr>
<tr>
<td>Female</td>
<td>436 (67.6)</td>
</tr>
<tr>
<td><strong>Race, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>442 (68.5)</td>
</tr>
<tr>
<td>African American</td>
<td>151 (23.4)</td>
</tr>
<tr>
<td>Others(^a)</td>
<td>52 (8.1)</td>
</tr>
<tr>
<td><strong>Ethnicity, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>17 (2.6)</td>
</tr>
<tr>
<td>Non-Hispanic</td>
<td>628 (97.4)</td>
</tr>
<tr>
<td><strong>Marital status, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Married or living with a partner</td>
<td>406 (62.9)</td>
</tr>
<tr>
<td>Not married(^b)</td>
<td>239 (37.1)</td>
</tr>
<tr>
<td><strong>Education, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>High school diploma or less</td>
<td>108 (16.8)</td>
</tr>
<tr>
<td>Some college or college degree</td>
<td>338 (52.4)</td>
</tr>
<tr>
<td>Graduate degree</td>
<td>199 (30.9)</td>
</tr>
<tr>
<td><strong>Monthly income (US $)</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;3000</td>
<td>241 (39.8)</td>
</tr>
<tr>
<td>3000-4999</td>
<td>168 (27.7)</td>
</tr>
<tr>
<td>≥5000</td>
<td>197 (32.5)</td>
</tr>
<tr>
<td><strong>Employment, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Employed full-time or part-time</td>
<td>372 (57.7)</td>
</tr>
<tr>
<td>Not employed(^c)</td>
<td>273 (42.3)</td>
</tr>
<tr>
<td><strong>Having chronic disease, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>558 (86.5)</td>
</tr>
<tr>
<td>No</td>
<td>87 (13.5)</td>
</tr>
<tr>
<td><strong>Number of chronic diseases</strong></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>2.5 (1.78)</td>
</tr>
<tr>
<td>Range</td>
<td>0.0-11.0</td>
</tr>
<tr>
<td><strong>Chronic disease (yes), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>High blood pressure</td>
<td>302 (46.8)</td>
</tr>
<tr>
<td>High cholesterol</td>
<td>241 (37.4)</td>
</tr>
<tr>
<td>Arthritis</td>
<td>221 (34.3)</td>
</tr>
<tr>
<td>Depression</td>
<td>184 (28.5)</td>
</tr>
<tr>
<td>Diabetes</td>
<td>144 (22.3)</td>
</tr>
<tr>
<td>Cancer</td>
<td>117 (18.1)</td>
</tr>
<tr>
<td>Kidney problems</td>
<td>84 (13)</td>
</tr>
<tr>
<td>Variables</td>
<td>Values</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Heart problems</td>
<td>79 (12.2)</td>
</tr>
<tr>
<td>Osteoporosis</td>
<td>63 (9.8)</td>
</tr>
</tbody>
</table>

**Internet usage hours per week**
- Mean (SD) 24.9 (20.78)
- Range 1.0-105.0

**eHealth literacy**
- Mean (SD) 31.2 (5.51)
- Range 8.0-40.0

**Patient portal use in the past 12 months, n (%)**
- 1 to 9 times 362 (52.7)
- About monthly 128 (18.6)
- More than monthly 197 (28.7)

**Perceived usability of the patient portal**
- Mean (SD) 36.6 (6.00)
- Range 6.0-42.0

**Concerns about privacy and security**
- Mean (SD) 2.7 (1.81)
- Range 1.0-7.0

---

**Beneficial Aspects of the Patient Portal**

A total of 734 comments were entered into the analysis of beneficial aspects of the patient portal. Table 2 summarizes the themes with a coded comment frequency greater than 5% of the total comment frequency. The most frequently mentioned aspect was communicating with health care teams using the eMessaging function (n=279, 38%). They favored being able to communicate quickly and directly with their care providers on nonurgent matters without having to make an appointment or a phone call:

*The ease of contacting my doctors. It eliminates the middle man and waiting that someone relays your question. I just send my doctor an email, and she responds right away.* [ID 235]
Another aspect that was most mentioned was about viewing test results and visit summaries (n=275, 37.5%). The participants also favored the aspect of managing appointments and receiving reminders (n=103, 14%). They stated that these features help them keep track of their upcoming schedules and health status:

- I like the ability to see my after-visit summaries and results, without having to call the provider’s office to try and track stuff down. [ID 719]
- What I like most is that it’s very convenient when scheduling appointments! And I love the fact that my test results are posted as soon as they come back! [ID 305]

Eighty-five (12%) comments indicated easy access to personal health information as an advantage, like the following comment:

- Availability 24 hours, ease of seeing information without need to bother office staff, able to print info to take with me to other doctors. [ID 373]

Other featured beneficial aspects included getting prescription refills and reviewing medications (n=63, 9%) and ease of use and convenience (n=45, 6%).

### Difficulties in Using the Patient Portal

Of 187 participants who answered that they experienced difficulties when using the patient portal, 179 of them specified the difficulties they had (Table 2). Difficulties in communicating with health care teams via eMessaging were mentioned the most (n=45, 25%), including not getting timely responses and not being able to send messages to providers they want:

- Some physicians and health providers don’t respond to patient message on the patient portal even don’t read the message and needs to call them and it waste a lot of time. [ID 848]
- Some providers do not use the email option and therefore makes it a bit more difficult to communicate with provider. [ID 429]

Difficulties related to the eMessaging function itself, such as character limit or file attachment, were also mentioned.

The next most frequently mentioned was difficulty in the log-in process (n=31, 17%). They stated difficulties such as entering passwords multiple times or taking a long time to obtain a verification code:

- At times I entered my password incorrectly and was contacted by the IT person and I was asked a lot of questions. Had to change my password. [ID 388]
- It takes 5 minutes to receive text for 3rd party verification to access the site, so by the time I get it, I’ve moved on to other things. I don’t have the time to sit around and wait to be able to enter a password. [ID 84]

Twenty-four (13%) comments were about difficulties in using the patient portal functions to set up appointments and refill medications:

- Scheduling appointments has been a challenge and the available listed is different than what they say on the phone. [ID 575]
- I tried to refill a prescription and the next day I couldn’t find any information as to my request being fulfilled. Had to call the doctor’s office to get refill completed. [ID 889]
Another notable difficulty was finding test results and other information in the portal (n=22, 12%):

- Tells me I have test results then there is no report. [ID 806]
- Looking through a lot of data to find out what I needed. [ID 1049]

About 8% of the comments mentioned that information or list of health care providers is not updated properly (n=15, 8%). Another 8% were about usability issues (n=14, 8%), including unclear display of information and difficulty navigating the functions. Issues with the patient portal system were also cited as difficulty (n=10, 6%), for example:

- The patient portal is always saying that it is deactivated. Then I must call to talk to a tech to help me get back on. [ID 919]

Table 3. Trust and concerns about privacy and security of the patient portal.

<table>
<thead>
<tr>
<th>Themes</th>
<th>Coded Comments, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Topic: Trust in privacy and security of your patient portal (n=554)</strong></td>
<td></td>
</tr>
<tr>
<td>Trust in health care system and network security</td>
<td>181 (33)</td>
</tr>
<tr>
<td>Protected registration and log-in process</td>
<td>173 (31)</td>
</tr>
<tr>
<td>Never been concerned or thought about privacy and security</td>
<td>107 (19)</td>
</tr>
<tr>
<td>Do not think that patient portal is fully safe and private but hope it is</td>
<td>57 (10)</td>
</tr>
<tr>
<td>HIPAA(^b) regulations, health care system’s privacy policies</td>
<td>40 (7)</td>
</tr>
<tr>
<td><strong>Topic: Concerns about privacy and security of your patient portal (n=122)</strong></td>
<td></td>
</tr>
<tr>
<td>Risk of data breaches (eg, hacking attacks and identity theft)</td>
<td>74 (61)</td>
</tr>
<tr>
<td>Personal information could be used against me by third parties</td>
<td>17 (14)</td>
</tr>
<tr>
<td>Distrust of the internet/computer/patient portal system</td>
<td>16 (13)</td>
</tr>
<tr>
<td>Risk of others accessing my personal information</td>
<td>14 (11.5)</td>
</tr>
</tbody>
</table>

\(^a\)Themes with a coded comment frequency greater than 5% of the total comment frequency are included.

\(^b\)HIPAA: Health Insurance Portability and Accountability Act.

Comments on security maintained by protected registration and log-in process were also frequently mentioned (n=173, 31%). The participants stated that personal verification is needed to sign up for the patient portal account, and personally owned information such as passwords or fingerprints is required to log-in. In particular, they appraised that the 2-factor authentication process strengthens the log-in security:

- I needed a password or fingerprint to access my file. [ID 539]
- Having a security code that hopefully only the patient would be able to access. [ID 377]
- It has two layers of security features in order to login. [ID 241]

About one-fifth of the comments mentioned that they had never been concerned or thought about the privacy and security of the patient portal (n=107, 19%). Many of them recognized that digital security is a web-wide concern, which cannot be completely guaranteed, for example:

- It didn’t concern me, because I knew the information was already online, or on computers capable of going online. Me having access to it doesn’t increase the risk of it being stolen unless I personally make a mistake. [ID 707]

Similarly, some participants commented that they do not think the patient portal is fully safe and private, but they hope it is (n=57, 10%):

- In this age and time nothing is 100% safe, but I am hoping they are using the right safeguards to protect my information. [ID 318]

Forty (7%) comments stated that federal statutes such as the HIPAA (Health Insurance Portability and Accountability Act) regulations and the health care system’s privacy policies established under applicable statutes ensure privacy and security:

- Bound by HIPAA regulations, so I considered the process to be secure. [ID 361]
Concerns About Privacy and Security of the Patient Portal

Of 126 participants who had a high level of concerns about privacy and security of the patient portal, 122 commented on reasons for their concerns (Table 3). The majority of the comments were related to the risk of data breaches (n=74, 61%). The participants expressed their concerns about hacking attacks and identity theft occurring in health care institutions and private companies:

- *Because of past breaches of health information at the health care system, as well as breaches with personal credit cards, consumer credit reporting agencies, etc.* [ID 822]
- *It’s private information and the demographic data maintained could easily be used to steal an identity.* [ID 322]

Similar to concerns about data breaches, 17 (14%) comments particularly mentioned that their personal information could be used against them by third parties. They shared their experiences of personal information being compromised by insurance companies or other types of business:

- *Hackers, insurance using it against me later etc.* [ID 228]
- *My data has been compromised multiple times. I have had at least four fraudulent credit cards taken out in my name. I feel my health info is just as vulnerable.* [ID 756]

Distrust of the internet/computer/patient portal system was mentioned in 16 comments (13%), like stating:

- *I’m always worried about Internet/information security. It’s one of the biggest issues of our day.* [ID 138]

The risk of others accessing my personal information on the internet was mentioned in 14 (11.5%) comments:

- *I’m always worried about my personal health information being in multiple places (creates extra opportunities for people to get ahold of it who shouldn’t have access).* [ID 116]

**Discussion**

**Principal Findings**

Our findings provide an empirical understanding of the lived experience of using a patient portal in adult patient users across care settings, focusing on the beneficial aspects and difficulties in using the patient portal, and trust and concerns about privacy and security. The majority of the participants perceived the patient portal functions as beneficial for communicating with health care teams and monitoring health status and care activities. At the same time, about a quarter of them shared specific difficulties they experienced while using those functions. Although the level of concerns about privacy and security was generally low among the participants, they provided practical feedback that the software company and health care system personnel could refer to in order to implement the patient portal more secure and private.

The beneficial aspects and difficulties in using the patient portal found in our study were fairly consistent with what we found in previous studies [2,8,15,27]. Interestingly, some participants perceived a certain patient portal function as beneficial, while others found it difficult to use the same function. Communicating with health care teams using the eMessaging function was the most preferred feature among the participants, but it was also the most mentioned difficulty. This is probably because the eMessaging function is one of the most frequently used patient portal functions by patients [28,29]. Our participants favored direct communication with care providers for nonurgent matters. On the other hand, there were participants expressing difficulties in not getting timely responses or not being able to send a message to the care providers they want. This finding emphasizes the importance of care providers’ involvement in using a patient portal as patients’ experiences of using the eMessaging function may largely depend on care providers’ use of that function [2,16].

Viewing test results and visit summaries was another aspect that was most mentioned as beneficial in terms of tracking health status, but about 12% of those who had difficulties in using the patient portal experienced difficulty finding such information in the portal. Different perceptions also coexisted for other patient portal functions such as refilling medications and scheduling appointments; there were participants who found these functions convenient and useful, while others felt that those functions need further improvement. These differences in perceptions could be attributed to each participant’s eHealth literacy, proficiency with HIT, and usability of the patient portal, which have been reported as factors associated with the adoption of patient portals [8-11]. Indeed, those who had difficulties in using the patient portal had lower mean scores of eHealth literacy (P>.39) and perceived usability (P<.001) than those who did not. Periodic evaluation of the usability by end users with different levels of eHealth literacy would help make the patient portal more user friendly. Timely updates of accurate information on patient portals by health care teams may also help mitigate difficulties in viewing medical records and using medication refill and appointment functions.

The log-in process was the second most frequently mentioned difficulty. Since only 4.2% (n=31) of our sample mentioned this difficulty, we may consider that the configuration of the patient portal satisfies overall ease of access, although there is still room for improvement. Similar to previous studies [8,30], there were participants pointing out the inconvenience of entering usernames and passwords every time they log in to the patient portal and having to contact the service desk when the account is locked after entering them incorrectly. They particularly commented that it often takes a long time to receive a separate code via email or text for 2-factor authentication at log-in. Ironically, this log-in process was recognized by about a quarter of our participants as a key factor in making them trust that privacy and security are maintained on the patient portal. Health care institutions and IT developers should focus on balancing the convenience of logging in desired by users with
the maintenance of security standards on the patient portal system.

Trust in the health care system and network security was most mentioned as what made the participants feel their information on the patient portal would be kept private and safe. Along with secure encrypted browsers, their faith in the health care institution and providers and the software company led to their trust in privacy and security of the patient portal. This aligns with the literature that a high level of trust in health care providers is an antecedent of fewer concerns about privacy and that providers’ encouragements positively influence each individual's acceptance and use of patient portals [31-33]. Of the 554 submitted comments on this topic, 7.2% of them mentioned HIPAA, which is a federal law enacted to protect individuals’ sensitive health information [34] and the health care system’s privacy regulations. Further research is recommended to assess the impact of raising knowledge and awareness of personal health information safeguards on patients’ trust in privacy and security of patient portals [33].

Of the 122 participants who commented on concerns about the privacy and security of the patient portal, about two-thirds of them mentioned the risk of data breaches. They were aware of hacking attacks reported by the media and shared their direct and indirect experiences of identity theft. They were particularly concerned about the potential risk of personal information being leaked and used by private entities such as insurance and credit card companies. Indeed, data breaches have become a serious threat in the health care sector. The number has increased steadily over the past decade; 4419 health care data breaches of 500 or more records have been reported in the United States between 2009 and 2021 [35]. There was another perspective that distrust of the internet system leads to concerns about privacy and security. However, about 22% of our participants did not think much about this matter, stating that nothing is completely safe on the internet. This rather fatalistic view seems similar to the fact that many people are aware of the potential risk of electronic financial transactions, yet accept the risk and use digital banking [32]. Health care institutions and software companies implementing patient portals should continuously monitor the privacy and security safeguards of patient portals and provide relevant information assurance when necessary.

Limitations
This study has several limitations. The survey recruited adult patients from a variety of 20 care settings from hospitals and clinics in a single health care system located in the mid-Atlantic region of the United States. The findings may not be generalizable since our sample and the patient portal (MyChart) included in this study cannot represent all patient portal users and programs. In addition, we only included those who had used the patient portal at least twice in the past 12 months prior to the survey; thus, infrequent users’ or nonusers’ perceptions of the patient portal were not reflected in this study. The relatively low survey response rate (7.5%) is another limitation that may affect the external validity of the study findings, although our response rate was similar to previous studies that used patient portal eMessages for participant recruitment [36,37].

Conclusions
This study investigated the lived experience of using a patient portal in adult patients recruited from multiple care settings in a large integrated health care system, focusing on their perceived benefits and difficulties in using the patient portal along with their perceptions on privacy and security. The findings showed that most participants recognized the convenience, ease of use, and usefulness of the patient portal functions for communicating with health care teams and tracking care activities. About a quarter of the participants shared their difficulties in using the patient portal functions in terms of eMessaging communication, log-in process, and finding information in the portal. While the participants’ concerns about the privacy and security of the patient portal were generally low, they provided insightful comments that could help health care institutions and software companies implement patient portals to be more secure and private. Future studies may consider targeting other patient portal programs and patients with infrequent or nonuse of patient portals.

Acknowledgments
This study was supported by the 2018 Small Grant from the Sigma Theta Tau International Honor Society of Nursing.

Conflicts of Interest
None declared.

References


Abbreviations

EHR: electronic health record
HIPAA: Health Insurance Portability and Accountability Act
HIT: health information technology

©Elisa H Son, Eun-Shim Nahm. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 25.07.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License
(https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Mixed Reality Technology to Deliver Psychological Interventions to Adolescents With Asthma: Qualitative Study Using the Theoretical Framework of Acceptability

Kelsey Sharrad¹, BSc; Caitlin Martini¹, BSc; Andrew Tai²,³, BSc, MD, PhD; Nicola Spurrier⁴, BSc, MD, PhD; Ross Smith⁵, BSc, PhD; Adrian Esterman⁶, BSc, PhD; Ian Gwilt⁷, BA, PhD; Debra Sandford⁸, BSc, PhD; Kristin Carson-Chahhoud¹,⁹,¹⁰, BSc, PhD

¹Translational Medicine and Technology Group, Australian Centre for Precision Health, University of South Australia, Adelaide, Australia
²Department of Respiratory and Sleep Medicine, Women's and Children's Hospital, Adelaide, Australia
³Robinson Research Institute, University of Adelaide, Adelaide, Australia
⁴Department of Health and Ageing, Government of South Australia, Adelaide, Australia
⁵Australian Research Centre for Interactive and Virtual Environments, University of South Australia, Adelaide, Australia
⁶UniSA Clinical and Health Sciences, University of South Australia, Adelaide, Australia
⁷UniSA Creative, University of South Australia, Adelaide, Australia
⁸Health and Medical Sciences Faculty, University of Adelaide, Adelaide, Australia
⁹Adelaide Medical School, University of Adelaide, Adelaide, Australia
¹⁰South Australian Health and Medical Research Institute, Adelaide, Australia

Corresponding Author:
Kelsey Sharrad, BSc
Translational Medicine and Technology Group
Australian Centre for Precision Health
University of South Australia
GPO Box 2471
Adelaide, 5001
Australia
Phone: 61 08 8302 2734
Email: Kelsey.hibberd@live.com.au

Abstract

Background: Interactive, mixed reality technologies such as augmented reality, virtual reality, and holographic technology may provide a novel solution to fast-track the translation of evidence into practice. They may also help overcome barriers to both mental health and asthma management service uptake, such as cost, availability of appointments, fear of judgment, and quality of care.

Objective: This study aimed to investigate if mixed reality technology is an acceptable mechanism for the delivery of a component of cognitive and behavioral therapies for the management of elevated psychological distress among young people with asthma.

Methods: To explore the perceived acceptability of these technologies, mixed reality tools were evaluated via qualitative, 1-on-1 interviews with young people with asthma and symptoms of psychological distress, parents/caregivers of young people with asthma and symptoms of psychological distress, and relevant health professionals. The Theoretical Framework of Acceptability was used for the deductive coding of the recorded interview transcripts.

Results: This study enrolled the following participants: (1) 3 adolescents with asthma and symptoms of psychological distress with a mean age of 14 (SD 1.7) years; (2) 4 parents/caregivers of adolescents with asthma with a mean age of 55 (SD 14.6) years; and (3) 6 health professionals with a mean age of 40.8 (SD 4.3) years. A total of 4 constructs—experienced affective attitude, experienced effectiveness, self-efficacy, and intervention coherence—were coded in all participant transcripts. The most frequently coded constructs were experienced affective attitude and intervention coherence, which were reported a total of 96 times. The least frequently coded construct was anticipated opportunity cost, which was reported a total of 5 times. Participants were mostly positive about the mixed reality resources. However, some concerns were raised regarding ethicality, particularly regarding...
privacy, accessibility, and messaging. Participants noted the need for technology to be used in conjunction with face-to-face engagement with health professionals and that some patients would respond to this type of delivery mechanism better than others.

**Conclusions:** These results suggest that mixed reality technology to deliver psychological interventions may be an acceptable addition to current health care practices for young people with asthma and symptoms of psychological distress.

**Trial Registration:** Australia and New Zealand Clinical Trials Registry ACTRN12620001109998; https://anzctr.org.au/Trial/Registration/TrialReview.aspx?id=380427

**(JMIR Hum Factors 2023;10:e34629) doi:10.2196/34629**

**KEYWORDS**

asthma; augmented reality; virtual reality; mixed reality; psychological distress; adolescent; cognitive behavioral therapies; mental health

**Introduction**

**Background**

Australia has one of the highest asthma prevalence rates in the world [1], with 11% of people being affected, as per self-report [2]. Asthma is the leading cause of disease burden among young Australians aged 5 to 14 years, with 460,000 (10%) young people affected [2].

According to a 2018 report, people with asthma were more likely to experience psychological distress (the experience of symptoms of anxiety and depression at subclinical levels [3]) than individuals without asthma (15% vs 8.7% for high levels and 11% vs 3.4% for very high levels, respectively) [4]. In adolescents, a 2014 survey of 533 Australians between 12 and 25 years of age found that half the adolescents with asthma experienced symptoms of heightened psychological distress [5]. The most common causes of distress were similar to those of adolescents without asthma; however, asthma-related problems also contributed to psychological distress in this sample. The study highlights that psychological distress is not uncommon and has both asthma and nonasthma–related triggers.

**Psychological Interventions for People With Asthma**

Due to the bidirectional relationship between asthma and symptoms of psychological distress [6], psychological interventions may offer techniques and strategies to manage both psychological distress and symptoms of asthma, thus reducing the risk of exacerbations [7]. Research in this population is limited, with a 2005 systematic review (currently being updated [8]) unable to draw conclusions about the potential role of psychological interventions for children with asthma due to heterogeneous data [9]. Cognitive and behavioral therapy (CBT) is a type of psychological intervention that helps patients recognize and modify thoughts and behaviors that may be detrimental to their health and well-being [10]. CBT-based strategies could be useful in the treatment of symptoms of heightened psychological distress in people with asthma; however, evidence suggests that engagement with treatment is low in this population, with reports estimating that 4 (80%) out of 5 children and adolescents who could potentially benefit from psychological intervention are reportedly not accessing it [11]. In a 2016 Australian survey, 48% of parents of adolescents aged between 12 and 17 years reported that their child refused help; however, 39% of parents were not sure where to get help, 33% could not afford help, and 29% reported that they could not get an appointment [12]. Even when services are accessed, a recent study conducted in 21 countries found that only 9.8% of individuals with an anxiety disorder received possibly adequate treatment based on evidence-backed guidelines [13].

**Technology for Health Interventions**

A recent systematic review of 28 studies found that text-based internet searches were the most commonly identified help-seeking approach among adolescents, along with other internet communities [14]. Reasons for this preference for help-seeking via the internet include anonymity and privacy, immediacy, ease of use, inclusivity, connection with others, and an increased sense of control [14]. However, the effectiveness and safety of unguided self-help vary significantly due to the wide range and quality of sources available [10]. Evidence-based psychological interventions delivered via technologies such as smartphone apps and online resources—also known as e-psychology, eHealth, or e-mental health interventions—have the potential to be an effective option for psychological well-being support and may increase access to and quality of care [15]. A 2016 meta-analysis exploring the use of digital CBT in children and adolescents reported reduced anxiety in the intervention group compared with the control. No statistically significant differences in efficacy were observed between digital and in-person treatment modalities [16].

**Mixed Reality for Health Interventions**

Interactive technologies such as augmented reality (AR), virtual reality (VR), and holographic technology (also known as mixed reality technology) may provide a novel solution to aid in the timely translation of evidence-based treatment into practice. Using a smartphone as a viewing device, AR superimposes digital information into the real world so that content seems to coexist with reality [17]. VR requires a headset to view content, allowing the viewer to feel completely immersed in the digital world. On the other hand, holographic technology uses the projection of diffracted light to create images. Mixed reality technologies deliver treatment and health care information through videos, graphics, and animation, which can address low health literacy [17,18], allow for tailoring for individual population characteristics (eg, age, language), increase engagement [19], increase accessibility of information [20], and enable real-time updates of content, thus reducing the evidence-to-practice gap. AR and VR have been studied in multiple health care contexts, including cancer, autism, and chronic pain, with feasibility, acceptability, and early efficacy
data suggesting that these modes of intervention delivery may be beneficial for adolescents [21-23].

In children and adolescents with asthma, a recent scoping review found that digital health interventions (including VR and AR interventions) were a promising option for asthma management and treatment delivery and were perceived positively by health care professionals and patients [24]. Methodologically rigorous research is needed to ensure that evidence-based, easily accessible digital interventions are made available [24,25].

Study Aims
Accordingly, this study aimed to investigate whether mixed reality technology is an acceptable mechanism for the delivery of a component of CBT to manage symptoms of elevated psychological distress among young people with asthma.

Methods
Primary Methodology
This paper focuses on the acceptability of mixed reality technology as a delivery mechanism for a component of CBT to manage symptoms of elevated psychological distress among young people with asthma. The mixed reality tools utilized for this qualitative study were preexisting resources, including 1 bespoke AR resource (Figure 1), 1 VR resource, and 1 holographic resource (Figure 2). For the AR resource, participants used a smartphone camera to make a digital image of a human body appear as though it were in real space (Figure 1). The digital image provides information about asthma and the effect it can have on various body systems, demonstrating how psychoeducation (a technique used in CBT) could be offered using this technology. For the VR resource, participants wore a cardboard headset over their eyes, into which a smartphone was placed. The smartphone played a VR-specific video of calming nature landscapes while a meditative audio track played. The holographic resources involved a small plastic projector (Figure 2) being placed over a smartphone. When a specific video was played, images appeared in 3D within the plastic projector. The VR and holographic resources demonstrate the capacity of these technologies to deliver mindfulness strategies. Three different technologies were chosen for hypothesis-generating purposes to establish if 1 mechanism is more acceptable than others.

Figure 1. This screenshot of the augmented reality (AR) technology shows a member of the research team sitting atop the paper resource. The AR has been activated by the smartphone camera.
To evaluate perceptions of usability and appropriateness, 1-on-1 interviews were carried out with the target audience. Semistructured moderator guides (Multimedia Appendices 1-3) were developed to direct interviews with adolescents with asthma, parents/caregivers of adolescents with asthma, and health professionals, including general pediatricians, psychologists, psychiatrists, nursing staff, and pediatric respiratory specialists. Participants were not provided with access to the mixed reality resources before the interview. During the interviews, the research staff conducted a brief training session, and participants were provided with the mixed reality resources with which to interact in their own time. Interviews ran for approximately 1 hour and took place in meeting spaces in the respiratory department at 2 large teaching hospitals in metropolitan Adelaide, South Australia, or online via Zoom software. Sessions were audio recorded, and verbatim transcripts were sent back to the participant for validation after the interview. Participants also completed a simple questionnaire (Multimedia Appendices 4-6) requesting demographic information, as well as self-reported measures of asthma knowledge and technology useability that will be explored in subsequent reports.

**Sampling/Recruitment**

**Overview**

Participants were all identified through a pediatric respiratory specialist at the Women’s and Children’s Hospital. Health professionals were recruited through word of mouth. Purposive sampling was utilized to ensure a good representation of participant characteristics to meet the requirements of the research question.

Hospital staff provided potential participants with a copy of the participant information sheet and consent form for their review. If the participants were interested in learning more, consent to contact them was obtained, and the research staff was given contact details to follow up for screening and consenting procedures.

**Inclusion Criteria for Young People**

Young people were eligible for inclusion in this study if they (1) were aged between 13 and 17 years, (2) were formally diagnosed with asthma by a health professional (inpatients or outpatients), (2) had experienced or were currently experiencing symptoms of psychological distress determined by the Kessler Psychological Distress Scale (K10+) [26], (3) had access to a smartphone with the owner’s permission to use it during the interview, and (4) were English speaking or able to understand written English.

**Inclusion Criteria for Parents/Caregivers**

Participants were eligible for inclusion in this study if they (1) were the parent/caregiver of an adolescent with asthma (aged 13 to 17 years) who currently had or had reported in the past elevated symptoms of psychological distress (did not need to be a child actively participating in this study), (2) had access to a smartphone and can use smartphone technology (basic level), and (3) were English speaking or able to understand written English.

**Inclusion Criteria for Health Professionals**

Health professionals were eligible for inclusion in this study if they (1) had been practicing in their respective fields for at least 12 months, (2) had access to a smartphone and could use smartphone technology (basic level), and (2) were English speaking and able to understand written English.

**Exclusion Criteria for All Participant Groups**

Participants with an intellectual disability or cognitive impairment that would inhibit their ability to provide informed consent and participate in the study were ineligible to participate. Young people with a history of epilepsy or other contraindications for the use of VR were also ineligible to participate.

**Qualitative Data Analysis**

Qualitative data were coded using three prespecified lenses to enable insight into different aspects of the mixed reality interventions: (1) the Theoretical Domains Framework (TDF)[2], the Theoretical Framework of Acceptability (TFA), and (3) the Enlight protocol. This paper will focus solely on data obtained through the TFA, while the TDF and Enlight protocols will be featured in separate reports. Deductive thematic coding was used with a framework analysis technique based on the TFA [27]. The TFA comprises 7 constructs reflecting the multifaceted nature of acceptability, incorporating both anticipated and experienced thoughts, beliefs, and feelings regarding the intervention [27,28]. The 7 constructs are ethicality, self-efficacy, intervention coherence, affective attitude, burden,
opportunity costs, and perceived effectiveness, with the last 4 separated into “anticipated” and “experienced” subcategories. The TFA was shown to be successful in exploring acceptability in health promotion interventions [29]. Previous research demonstrates a more robust understanding of acceptability when a framework is applied compared with no framework [27-29]. All transcripts were coded by 2 independent researchers (authors KS and CM), with discrepancies resolved through consensus or discussion with a third party (author KCC). During the coding process, quotes were determined to be generally positive, negative, or neutral toward the mixed reality technology. A standardized pilot-tested data extraction Microsoft Excel (Microsoft Corp) template was used for data management.

Ethics Approval
This study was conducted following the principles of the Declaration of Helsinki [30] and received ethical approval from the Human Research Ethics Committee for the Women’s and Children’s Health Network (HREC/18/WCHN/172) and the University of South Australia Ethics Committee (201967).

Table 1. Participant demographic data.

<table>
<thead>
<tr>
<th>Participant group</th>
<th>Young people with asthma (n=3)</th>
<th>Caregivers of young people with asthma (n=4)</th>
<th>Health professionals (n=6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>14 (1.7)</td>
<td>55 (14.6)</td>
<td>41 (4.3)</td>
</tr>
<tr>
<td>Female sex, n (%)</td>
<td>2 (66)</td>
<td>1 (25)</td>
<td>4 (66)</td>
</tr>
<tr>
<td>Nationality, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td>3 (100)</td>
<td>4 (100)</td>
<td>5 (83)</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>N/A\textsuperscript{a}</td>
<td>N/A</td>
<td>1\textsuperscript{b} (17)</td>
</tr>
</tbody>
</table>

\textsuperscript{a}N/A: not applicable.
\textsuperscript{b}Participant was of dual UK and Australian nationality.

Qualitative Analysis

Overview
A total of 4 constructs, namely, experienced affective attitude, experienced effectiveness, self-efficacy, and intervention coherence, were coded in all 13 participant transcripts. The most frequently coded constructs were experienced affective attitude and intervention coherence, which were reported a total of 96 times, while the least frequently coded construct was anticipated opportunity cost, which was reported a total of 5 times. The remaining categories were coded between 11 and 85 times. Example quotes for each TFA construct are included in Multimedia Appendix 7.

Participant groups were similar in their proportion of positive, negative, and neutral quotes across all constructs (Table 2).

Table 2. Positive, negative, and neutral quotes obtained from participant interviews and organized by participant group.

<table>
<thead>
<tr>
<th>Participant group</th>
<th>Positive, n (%)</th>
<th>Neutral, n (%)</th>
<th>Negative, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health professionals</td>
<td>138 (56)</td>
<td>78 (32)</td>
<td>29 (12)</td>
</tr>
<tr>
<td>Young people with asthma</td>
<td>48 (64)</td>
<td>23 (31)</td>
<td>4 (5)</td>
</tr>
<tr>
<td>Parents of young people with asthma</td>
<td>79 (66)</td>
<td>32 (27)</td>
<td>8 (7)</td>
</tr>
</tbody>
</table>

Affective Attitude
Affective attitude is defined as “how an individual feels about taking part in an intervention” [27]. Transcripts were coded with quotes before and after exposure to the intervention to assess anticipated and experienced affective attitudes. There were 25 quotes identified for anticipated affective attitude, with most (n=13, 52%) being positive. Most quotes were identified in interviews with health professionals, with a common theme emerging of the potential for technology to be a useful tool when used in conjunction with a face-to-face consultation with physicians.

Results

Participants
A total of 19 participants were approached to take part in the study, with 13 completing interviews. Two young people and 1 parent withdrew from the project for psychological reasons, and another 2 young people and 1 parent were considered ineligible. Interviews were conducted by 1 of 3 researchers (authors KS, ZK, and KCC).

The 13 participants included 3 (23%) adolescents with asthma, 4 (31%) parents of young people with asthma, and 6 (46%) health professionals, including 2 (33%) psychologists, 1 (16%) psychiatrist, 1 (16%) medical consultant, 1 (16%) respiratory nurse consultant, and 1 (16%) respiratory sleep nurse consultant. The average duration of experience among health professionals was 17.08 (SD 5.87) years. As per the inclusion criteria, all adolescents reported to have experienced or were currently experiencing symptoms of psychological distress determined by the K10+ scale. The technology experience level of participants varied greatly. Full demographic details are presented in Table 1.
I think on its own, I would be worried, but if they're also seeing a therapist, counselor, psychologist...or had a trusted adult and health professional to talk to, I think it could be very positive. [Health professional #6]

Young people were also generally positive about the idea of technology:

Um, yeah, definitely. It would help. A lot because you know a lot of people spend most of their time on their phones nowadays, having an app like that. Well, you know, it would probably help a lot and you would be more, you know, you would probably use it more. [Young person #7]

Similarly, 96 quotes were identified for experienced affective attitude, with the majority (n=73, 76%) being positive. “Calming,” “cool,” and “engaging” were terms often used by participants in all groups to describe their experience. However, not all experiences were positive. One health professional suggested:

Uh, I probably couldn’t be bothered doing it myself. I guess. Um, I think sometimes it’s just about grabbing people’s attention and to me, I felt that had a sense that young people would say, yeah, yeah, whatever. [Health professional #5]

The comment that the holographic technology did not meet expectations commonly occurred among health professionals and parents of young people, along with disappointment about the size of the hologram. As 1 health professional stated:

I think if you kept, like, if you kept it, that kind of size, I think you’d lose engagement. Cause it would just be too little. But if you could have that, so it was, like, a really big thing then that’d be cool. [Health professional #2]

**Perceived Effectiveness**

Effectiveness refers to “the extent to which the intervention is perceived as likely to achieve its purpose” [27]. Once again, transcripts were coded to assess both anticipated and experienced effectiveness. A total of 23 quotes were identified for anticipated effectiveness, among which 11 (48%) were positive, 2 (9%) were negative, and 9 (39%) were neutral. When asked if this kind of technology could be useful for young people with asthma, young people tended to answer in the affirmative, with 1 participant explaining:

Um, yeah, definitely. It would help. A lot because you know a lot of people spend most of their time on their phones nowadays, having an app like that. Well, you know, it would probably help a lot and you would be more, you know, you would probably use it more. [Young person #7]

As with affective attitude, quotes coded for this domain in transcripts with health professionals focused on the need for technology to be used in conjunction with face-to-face engagement with health professionals. Participants also noted that some people would respond well to information delivered via technology, and others would not. One health professional said:

I think it’s a really good adjunct because there are some people that will be, like, perfect for... And there are some people that it will be okay with and there’ll be some people that won’t engage with it or won’t have the actual technology resources themselves, like the phone and the iPads... have access to do it. So, I think it just adds another string to your bow; it’s some other way that will work. And it’s about figuring out that goodness of fit. [Health professional #4]

A total of 85 quotes were coded for experienced effectiveness, among which 56 (66%) were positive, 8 (35%) were negative, and 21 (25%) were neutral. Participants felt that the technologies would be beneficial, particularly for education, given the engaging visual nature of the information. As 1 parent of a young person with asthma stated:

So, the goal is to understand as best you can, what’s happening either inside your own body or inside your child’s body. Okay. So that’s, that’s the aim. What is the best mechanism for doing that? Well, to see it really, isn’t it? And so then to see it in the most representative and the most real and the most, um, sort of engaging way becomes I guess the goal. [Parent of young person #2]

Once again, participants in all groups suggested that some people would respond more positively than others. A young person with asthma said:

It would be very beneficial for them. At least, um, or at least some people, depending on if they prefer it this way or that way. It would be very beneficial to the people who absolutely prefer something like this to learn and to discover more about themselves. [Young person #3]

Participants also noted the ease of accessibility of the information. One health professional stated:

Yeah, definitely because that’s something that they can access at home and it’s something that they can access at any time. Um, it’s creating that, um, autonomy to the, to the patient, to the family, um, re...reinforcing education that they might want, but in their own time and privacy of their own home. [Health professional #3]

**Ethicality**

Ethicality is “the extent to which the intervention has good fit with an individual’s value system” [27]. A total of 15 quotes were coded for this construct, with the majority (n=8, 53%) being classified as neutral, 2 (13%) as positive, and 5 (33%) as negative. Of these quotes, 12 (80%) were coded in interviews conducted with health professionals, and none were from interviews with young people. Issues identified included questions of accessibility:

They might not have access to reliable internet, or they may not have the money or access to a device, then, you know, it’s all very well having that, but they
Opportunity Costs

Opportunity costs refer to “the extent to which benefits, profits, or values must be given up to engage in the intervention” [27]. Transcripts were coded to assess both anticipated and experienced effectiveness. A total of 5 quotes were coded for anticipated opportunity costs, all from interviews with health professionals. Health professionals were asked if they thought that learning to use this kind of technology would be a good use of their time; 4 (80%) quotes were positive, and 1 (20%) was negative. The positive comments were replies to the question in the affirmative (“Yeah. Yes. Because then it would help me to keep up with the young people and know what they’re talking about”), while the negative comment expressed that the technology would not be relevant to her work (“Not for me. No.”)

In total, 11 quotes were coded for experienced opportunity costs, with quotes identified from all 3 participant categories of participant. Among these, 8 (73%) were positive, and 1 (9%) was neutral. Health professionals commented on whether learning to use and recommend the technology would be a good use of their time:

Yeah, definitely because that’s something that they can access at home, and it’s something that they can access at any time. [Health professional #3]

They also commented on the accessibility of the technology for intervention delivery:

Particularly if it’s, say, if it’s an app you download on your phone, if it’s a card box that you can easily access, think what people are looking for is that convenience, just download the app. That’s sort of a low-cost, no-cost app or the little cardboard boxes. And that again in my sort of world in education. They’re easy recommendations to make. [Health professional #4]

Likewise, young people with asthma and parents of young people with asthma commented predominantly on the accessibility of the technology. One parent noted that the accessibility would depend on the skills and abilities of the user:

Um, I would say that if it’s on the phone, it’s more accessible, probably quicker, but that would also depend on your knowledge [inaudible] accessing it. [Parent of young person #6]

Burden

This construct was also separated into anticipated and experienced subcategories, and it is defined as “the perceived amount of effort that is required to participate in the intervention” [27]. A total of 14 quotes were identified for anticipated burden, with 6 (43%) being positive, 1 (7%) being negative, and 7 (50%) being neutral. Most comments about anticipated burden related to the age of the participants, with comments identified in interviews with all 3 groups of participants. When asked if they thought the technologies would be difficult to use, a health professional said, “For me? Yes. For a young person? No.” One parent of a young person agreed:

Probably difficult for my age group. Um, because you know, we didn’t grow up with them, but the kids it’s just like, they don’t find it difficult at all. [Parent of young person #5]

Young people with asthma also echoed that sentiment (“Um, in some cases, maybe”), while another parent of a young person with asthma disagreed:

I think the awareness of technology these days of people my age as well as the younger generation is pretty reasonable. So, most things you pick up on pretty quickly. [Parent of young person #2]

Similarly, some health professionals expressed confidence in their ability to use the technology:

Um, no. I think everyone’s pretty up to date relatively with smartphones and that these days, so… no, not really difficult. [Health professional #3]

However, another expressed concerns:

I’ve noticed that I’m getting to an age now where I might not be quite up to speed with all, all the technology. So, you know, that can be a barrier as well if the clinicians and the treating teams aren’t up to speed. [Health professional #3]

A total of 46 quotes were coded for experienced burden, with the majority (n=30, 65%) being positive. Additionally, 7 (15%) comments were negative, and 5 (11%) were neutral. Many participants used the words “easy” and “simple” to describe their experience using the technologies, expressing that it would be easy to use once familiar with the technologies. As 1 parent of a young person with asthma asserted, “Um, I suppose anything like that is easy once you’re using it for a while.” Once again, age was cited as a confounding factor, with a young person with asthma saying, “It would be] probably easy for young people, and it’d be harder for older people.” Notably, participants expressed concerns about the burden of holographic technology more than the AR and VR tools. One parent of a young person with asthma said:

I think we’d have to say it was moderately difficult. Right? If I didn’t have a skilled proponent of the technology…I would have seen nothing cause you,
but you could imagine a little arrangement that has a black bit of cardboard at the back and then something that sort of positions that somehow, need a little bit of thought...and kind of got to position it from above, don't you, otherwise you interrupt one of your pictures. [Parent of young person #2]

Self-efficacy
This construct refers to “the participant’s confidence that they can perform the behavior(s) required to participate in the intervention” [27]. A total of 29 quotes were coded for this construct, with 17 (59%) being positive, 5 (17%) being negative, and 6 (21%) being neutral. Similarly, in the quotes identified for the burden construct, participants often shared that they found the technology difficult but that a younger person probably would not. One health professional explained “I’m not tech savvy,” and I parent shared that while some people of his age would adapt well to technology, he found it more difficult:

Look, uh, certainly there would be a certain amount in my generation. I would think that um, possibly, a little bit more, uh... You know, adaptable to computers and things. There’s some pretty smart people out there of my age that, you know, it’s. Computers and that, and technology is natural for them. It’s never been sort of part of what I’ve ever done if you know what I mean? [Parent of young person #6]

Likewise, when the young people with asthma were asked if they would use the tools themselves, most answered in the affirmative, and most indicated that they were comfortable using the technology:

Yes of course I would. These like, really work, very informative and very, very calming. [Young person #7]

Intervention Coherence
Intervention coherence is defined as “the extent to which the participant understands the intervention and how it works” [27]. There were 96 quotes coded for this construct. While some were unsure about the technology before experiencing the intervention (“I have heard of it. I don’t sort of understand it too much”), most demonstrated understanding afterward. We also considered that quotes about how the technology could be used outside of this project demonstrated an understanding of the intervention and how it worked and thus were appropriate to this construct. For example, 1 health professional said:

Yeah. To be used as, like, a treatment tool for those kids? I guess it depends on what you’re actually wanting to deliver, but I can definitely see [the] potential that if you’re giving, um, if you’re giving education on how like lungs work and how the pathophysiology of it all and the effectiveness and how medications work like am, your bronchodilators or whatever, that it could be really good cause I can actually see what’s happening inside of them. So, I think that would then help them kind of put the picture together as opposed to just talking to them.

From an anxiety and depression point of view. Um, yeah, I mean, I think that like a guided meditation type thing, like you did with the [VR] one that could, that could be useful in that kind of scenario. [Health professional #2]

Discussion
Principal Findings
The findings of this study suggest that mixed reality technologies are generally acceptable to adolescents with asthma, parents of adolescents with asthma, and health professionals. Participants across all 3 groups largely felt positively toward the mixed reality technology (Table 2) and considered the technology to be potentially effective and easy to use. This is also seen in a recent study in hospitalized children and adolescents with chronic illnesses, which found that, when compared to face-to-face CBT, a VR treatment involving education, breathwork, and mindfulness techniques received higher perceived efficacy scores [22]. Similarly, a 2020 study [23] exploring the feasibility and acceptability of a VR intervention for psychological well-being in children and adolescents with cancer found that the technology was viewed positively by health professionals, parents/caregivers, and patients. The largely positive affective attitude toward the mixed reality resources is promising, as affective attitude is an important determinant of behavior [31] and has been demonstrated to be a predictor of engagement with health behaviors [32-35]. Many participants—particularly health professionals—highlighted the need for technological interventions to be offered as an adjunct to existing resources. They also stressed the importance of face-to-face communication with care providers to ensure patient understanding of treatment instructions and maintenance of ongoing relationships between patients and health care professionals. This is supported by previous literature, with authors suggesting that technology-delivered health care may be suitable to provide support between face-to-face appointments [24]. It was also noted that patients respond differently to different modes of treatment delivery, with some preferring in-person communication and others preferring online modes. Health professionals were also most likely to have concerns relating to the opportunity cost of the intervention (eg, time taken to learn the technology and teach patients). This fits with recent literature, which states that health professionals are reported to be time-poor and overburdened [36].

Concerns were raised among all 3 groups of participants surrounding the ethicality of the intervention (ie, how well the intervention fit with the participants’ values), but this was mostly discussed by health professionals. Privacy and accuracy of messaging were identified as potential issues, as well as the accessibility of technological interventions for all socioeconomic groups. Issues surrounding privacy and mobile technologies in health care have already been identified in prior research, with some participants expressing concerns about the safety of medical data collected or transmitted via mobile devices [37,38]. Accessibility issues are also of legitimate concern; while 9/10 Australians own a smartphone, rates of use are still low in some groups, including those with disabilities and low household
income [39]. Additional research is warranted to explore the potential effects of this disparity.

Interestingly, multiple participants—most commonly health professionals and parents of adolescents with asthma—reported being disappointed with the holographic technology due to high expectations based on the portrayal of holographic technology in movies and TV shows. These preconceived perceptions may have introduced bias in the assessment of this type of technology. Future research with a larger budget may consider upsampling this type of resource to meet preconceived perceptions. Participants also expressed concerns about the burden of using holographic technology more so than VR or AR. The feedback comparing the mixed reality technology modalities led us to amend protocols for future studies limiting intervention content to AR.

**Limitations**

The sample of participants was all recruited from 1 hospital in Adelaide, South Australia, limiting the generalizability of the results to the larger population. While smaller than originally planned, this sample size was considered appropriate for the project timeline. Sim et al [40] included such “rules of thumb” as “between 12 and 20 participants in interview studies,” “2 to 10 participants in order to achieve redundancy or saturation,” and “at least five 1-hour interviews for theoretical saturation in grounded theory studies.” Furthermore, recent qualitative studies report similar sample sizes [41-45].

**Implications for Practice**

Findings from this study demonstrate that mixed reality resources may be an acceptable treatment/intervention delivery mechanism for young people with asthma. Participants reported feeling positive about the technology and the potential efficacy of this delivery mechanism, and the technology was largely considered easy to use. Health care services might consider the use of mixed reality technology in conjunction with existing resources to diversify treatment delivery and increase engagement.

**Conclusion**

The results of this study suggest that mixed reality resources may be an acceptable addition to current health care practices for the purpose of delivering psychological interventions to young people with asthma. Participants were mostly positive about the mixed reality resources; however, some concerns were raised regarding the ethicality, particularly in relation to privacy, accessibility, and accuracy of messaging. Participants noted the need for technology to be used in conjunction with face-to-face engagement with health professionals and noted that some patients would respond to this type of delivery mechanism better than others. Further randomized trials are warranted to explore the effect of mixed reality resources on health and behavior outcomes in this population.

**Acknowledgments**

This project was funded by the Channel 7 Children's Research Foundation, the University of South Australia Cancer Research Institute, and author KCC’s National Health and Medical Research Council (NHMRC) Translating Research Into Practice (TRIP) fellowship, along with discretionary funds supplied by the university to support the author’s fellowship. We would like to acknowledge and thank all participants for agreeing to take part in this study, as well as the staff at the Women’s and Children’s Hospital for facilitating recruitment.

**Conflicts of Interest**

None declared.

**Multimedia Appendix 1**

Moderator guide for interviews with young people with asthma.
[DOCX File, 17 KB - humanfactors_v10i1e34629_app1.docx ]

**Multimedia Appendix 2**

Moderator guide for interviews with parents of young people with asthma.
[DOCX File, 17 KB - humanfactors_v10i1e34629_app2.docx ]

**Multimedia Appendix 3**

Moderator guide for interviews with health professionals.
[DOCX File, 22 KB - humanfactors_v10i1e34629_app3.docx ]

**Multimedia Appendix 4**

Questionnaire for young people with asthma.
[DOCX File, 31 KB - humanfactors_v10i1e34629_app4.docx ]

**Multimedia Appendix 5**
Questionnaire for parents of young people with asthma.
[DOCX File , 30 KB - humanfactors_v10i1e34629_app5.docx ]

Multimedia Appendix 6
Questionnaire for health professionals.
[DOC File , 70 KB - humanfactors_v10i1e34629_app6.doc ]

Multimedia Appendix 7
Supporting quotes from interviews.
[DOCX File , 21 KB - humanfactors_v10i1e34629_app7.docx ]

References


Abbreviations

AR: augmented reality
CBT: cognitive and behavioral therapy
K10+: Kessler Psychological Distress Scale
TDF: Theoretical Domains Framework
TFA: Theoretical Framework of Acceptability
VR: virtual reality

Edited by A Kushniruk; submitted 01.11.21; peer-reviewed by AL Gainet, B Paladi, B Chaudhry, H Ayatollahi; comments to author 20.01.22; revised version received 10.08.22; accepted 10.11.22; published 26.07.23.

Please cite as:
Mixed Reality Technology to Deliver Psychological Interventions to Adolescents With Asthma: Qualitative Study Using the Theoretical Framework of Acceptability
JMIR Hum Factors 2023;10:e34629
URL: https://humanfactors.jmir.org/2023/1/e34629
doi:10.2196/34629
PMID:37494096

©Kelsey Sharrad, Caitlin Martini, Andrew Tai, Nicola Spurrier, Ross Smith, Adrian Esterman, Ian Gwilt, Debra Sandford, Kristin Carson-Chahhoud. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 26.07.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Original Paper

User-Centered Design of a Digitally Enabled Care Pathway in a Large Health System: Qualitative Interview Study

Maggie McCue¹, MS, RD; Rasha Khatib², MHS, PhD; Christopher Kabir², MSPH; Chris Blair¹, BA, MS, CCRP; Ben Fehnert³, MA; James King⁴, MA; Alexander Spalding⁴, MPhil; Lara Zaki⁵, MSc; Lambros Chrones¹, MD; Anit Roy¹, BS; David E Kemp⁵, MD, MS

¹Takeda Pharmaceuticals U.S.A., Inc., Lexington, MA, United States
²Advocate Aurora Research Institute, Advocate Health Care, Downers Grove, IL, United States
³Cognition Kit, Cambridge, United Kingdom
⁴Ctrl Group / Fora Health, London, United Kingdom
⁵Advocate Aurora Health, Downers Grove, IL, United States

Corresponding Author:
Chris Blair, BA, MS, CCRP
Takeda Pharmaceuticals U.S.A., Inc.
95 Hayden Ave
Lexington, MA, 02421
United States
Phone: 1 708 519 0271
Email: Chris.Blair@takeda.com

Abstract

Background: Major depressive disorder (MDD) is a leading cause of disability worldwide. Management of chronic conditions such as MDD can be improved by enhanced patient engagement, measurement-based care (MBC), and shared decision-making (SDM). A user-centered design approach can improve the understanding of the patient journey and care team workflows and thus aid the development of digital health care innovations optimized for the needs of patients living with MDD and their primary care teams.

Objective: This study aims to use qualitative research methods for the user-centered design of a digitally enabled MDD care platform, PathwayPlatform, intended to enhance patient engagement, MBC, and SDM.

Methods: Insights were gathered through 2 stages of qualitative interviews by a study team with expertise in qualitative research and user-centered design methods. Thematic analysis was used to generate an overarching understanding of a set of shared experiences, thoughts, or behaviors across a broad qualitative data set, including transcripts of interviews, to allow both inductive and deductive insights to emerge. Thematic analysis of interviews was supported by Dedoose (SocioCultural Research Consultants, LLC), a qualitative data analysis software tool that enables systematized coding. Findings and insights were presented based on code frequency, salience, and relevance to the research project.

Results: In stage 1, interviews were conducted with 20 patients living with MDD and 15 health care providers from September 2018 to January 2019 to understand the experiences with and perceptions about the initial functionality of the Pathway app while also exploring the perceptions about potential additional features and functionality. Feedback about care team workflows and treatment approaches was collected in stage-2 interviews with 36 health care providers at 8 primary care sites. Inductive and deductive thematic analyses revealed several themes related to app functionality, patient-provider engagement, workflow integration, and patient education. Both patients and their care teams perceived the remote tracking of patient-reported outcomes via digital tools to be clinically useful and reliable and to promote MBC and SDM. However, there was emphasis on the need to enhance the flow of real-time data shared with the care team, improve trend visualizations, and integrate the data within the existing clinical workflow and educational programs for patients and their care teams. User feedback was incorporated into the iterative development of the Pathway app.

Conclusions: Ongoing communication with patients living with MDD and their care teams provided an opportunity for user-centric developmental iterations of the Pathway Platform. Key insights led to further development of the patient-facing and care team–facing visit preparation features, collaborative goal-setting and goal-tracking features, patient-reported outcome
summaries, and trend visualizations. The result is an enhanced digital platform with the potential to improve treatment outcomes and provide patients living with MDD additional support throughout their treatment journey.

(JMIR Hum Factors 2023;10:e42768) doi:10.2196/42768

KEYWORDS
depression; major depressive disorder; depression management; patient engagement; user-centered design; mobile app; digital platform; qualitative research; shared decision-making; measurement-based care; mobile phone

Introduction

Background
Depression is a leading cause of disability worldwide, affecting nearly 300 million people [1,2]. Major depressive disorder (MDD) is a growing problem in the United States, with the total number of US adults with MDD increasing by 12.9%, from 15.5 million to 17.5 million between 2010 and 2018, and it is associated with a significant economic burden [3]. Primary care centers are the largest mental health service providers for people living with MDD, with up to two-thirds of visits to health care providers (HCPs) for depression occurring in a primary care setting [3-6]. Time constraints and the need to frequently manage multiple conditions during a single visit to a primary care setting can make it difficult for HCPs to fully engage with patients when it comes to their treatment for MDD [4]. Frequent communication and engagement between people living with MDD and their care teams may improve therapeutic outcomes, especially for chronic conditions such as MDD [4].

Measurement-Based Care
The American Psychiatric Association clinical guidelines recommend measurement-based care (MBC) for treating depression. MBC includes the routine use of standardized outcome measures to assess changes in depression symptoms, level of functioning, and quality of life across the treatment course [7]. MBC for MDD is effective because it allows the primary care provider to quantify clinical outcomes. This provides guidance for timely treatment modifications that may better meet the needs of the person being treated [8]. Furthermore, treatment decisions based on MBC give people living with MDD a better understanding of how their condition is changing over time and therefore potentially empowers them in terms of their own care [8]. In addition, compared with usual care, MBC in the management of depression has been shown to improve treatment adherence, thereby leading to improvements in clinical outcomes [8,9]. Despite the demonstrated benefits of MBC in improving treatment outcomes and patient engagement, adoption of MBC in routine clinical practice has been slow, with only 20% of HCPs using it in their practice [8]. Increased consultation times are often cited as a barrier to MBC implementation.

Shared Decision-Making
Along with MBC, another critical factor shown to improve treatment outcomes is the involvement of people living with MDD in the decision-making processes of their treatment journey [10]. Several studies have found that people experiencing a mental illness want to play a large role in the treatment decision-making process [10]. A shared decision-making (SDM) model of interaction can foster patient-provider engagement by empowering patients to play a great role in the decision-making process, thus creating an opportunity for them to have their voices, beliefs, values, goals, experiences, and preferences reflected in the treatment planning and monitoring process. This, in turn, can help increase treatment satisfaction and overall treatment adherence [10].

Among people receiving treatment for mental health disorders (including depression), increased adherence to treatments (including both psychopharmaceutical and psychotherapeutic) has been reported when they recognize that their treatments reflect their unique needs and preferences. A strong alignment between the treatment goals of the person living with depression and their treatment provider is another critical factor shown to be important in promoting adherence [10,11]. Several studies have reported misalignment between what patients and their treatment providers consider to be the most important treatment goal [10]. Thus, involvement in SDM is associated with a high probability of receiving quality care and improvement in symptoms [12] through increased adherence to drug treatment [10].

Digital Tool Development in MDD
Digital tool development that enhances patient engagement, MBC, and SDM has the potential to improve treatment outcomes in MDD [13]. Digital communications and information technologies have previously been shown to improve health care delivery by improving communication between providers, and by decreasing the need for face-to-face appointments, thus helping to alleviate the workload of HCPs [14,15]. With rapid advances and the adoption of smartphone technology, mobile health apps have generated interest from both the public and medical communities [16]. For the HCP, digital technology platforms, such as mobile apps, can offer low-cost interventions to monitor and improve services for patient populations that are difficult to retain during treatment [15]. A mobile app could help save time because people living with MDD could engage with symptom assessments (such as the 9-item Patient Health Questionnaire [PHQ-9]) and other clinical instruments outside their visit, whereas HCPs would only need to review these results instead of administering the instrument during the visit itself [8]. A mobile app could therefore promote MBC by allowing patients to remotely engage with validated quantitative measures of assessments, which can then be uploaded into their electronic health records (EHRs) for in-office visits and physician monitoring [8]. Collection of PHQ-9 results through a mobile app has been shown to be as sensitive as, or even more sensitive than, the traditional (in-person and paper-based) PHQ-9 data collection method [17]. This may be because people living with MDD might feel more comfortable revealing their
symptoms in remote settings through a mobile app rather than in traditional in-person settings.

From the patient’s perspective, several studies using smartphone app–based interventions for depressive disorders have shown that depressive symptoms were reduced significantly more with smartphone apps [16,18,19] through motivating some users to consult medical professionals for diagnosis and management [20,21] and through self-observation [22,23]. Furthermore, apps dedicated to the caregivers of people living with MDD have also been shown to help caregivers in better supporting their loved ones and to destigmatize mental health care [24].

User-Centered Design Approach
Achieving this potential of digital tools to improve treatment outcomes requires a nuanced understanding of the desirability and usability of patient-facing and care team–facing digital interfaces alongside the practical requirements for patient and care team adoption. Perceived utility and overall value of the product to care teams and privacy and confidentiality concerns are often cited by care teams as examples of barriers that can limit the adoption of these types of apps [25]. Various mobile apps are available for depression management; however, many are patient-facing only and do not include a care team interface [26]. The lack of guidance or feedback from the care team has often been cited by people living with depression as a barrier to the adoption of many apps that are dedicated to mental health care [25,27,28]. Patients might, for example, perceive the lack of feedback or engagement from the care team as indicating that the information they are providing through the app is not being monitored or integrated into their care processes, thus disincentivizing their engagement with these products [29].

A user-centered design approach can improve the understanding of the patient journey (Multimedia Appendix 1) [30]. Patient insights help develop an understanding of user needs, with iterative designs and prototypes playing a key role in how these insights and needs can be unearthed [30]. However, studies describing the user-centered design approach in health-related technology remain limited [30]. Takeda, Lundbeck, and Advocate Aurora Health (AAH) partnered to cocreate a digitally enabled care experience with users (patients living with MDD and care teams), software developers, and health-technology product development specialists (Ctrl Group and Fora Health) [31]. Specifically, a digital mobile patient interface, the Pathway App, was designed with a conversational interface and tested via a pilot feasibility study with 40 patients living with MDD [31]. The study showed a trend toward high patient activation and patient-provider engagement for people who used the app in addition to usual care compared with those who were assigned to usual care only [31]. Building on these results, a new iteration of the app was created and incorporated into a new digital platform, Pathway Platform, which includes the Pathway App; EHR-integrated, real-time, patient-level data sharing; and educational programming that is both care team and patient facing. This was guided by a more comprehensive understanding of care team workflows and patient and care team insights [32].

In this paper, we describe how user-centered design was applied to develop a digitally enabled MDD care platform that is optimized for the needs of patients living with MDD and their primary care teams.

Methods
Overview
This report describes the qualitative research undertaken to understand, iterate, and integrate Pathway Platform into primary care in the AAH system within and around Chicago, Illinois. In stage 1, interviews were conducted with 20 patients with MDD who participated in the pilot feasibility study and 15 HCPs (from September 19, 2018, to January 30, 2019). The stage-1 interview sought to understand the experiences with and perceptions about the initial functionality of the Pathway app (Multimedia Appendix 2) while also exploring perceptions about potential additional features such as HCP visit preparation, patient education, and goal setting and tracking. Interviews with patients lasted up to 60 minutes, and interviews with HCPs lasted up to 30 minutes. Another round of follow-up feedback was collected from 36 HCPs at 8 primary care sites through stage-2 interviews to understand care team workflows in the treatment of patients with MDD, the extent to which Pathway Platform can help optimize care, and what support the care team will need to make Pathway Platform work at their respective sites. HCPs were included if they were involved in primary care and specifically managed the care of patients with MDD. The semistructured, in-person interviews were conducted by a study team with expertise in qualitative research, user-centered design methods, workflow assessment, and educational support (Ctrl Group and PRIME Education LLC). Care teams were interviewed about topics related to workflow, perceptions about and experiences with MBC and SDM, and educational needs of patients with MDD and their care team members. The interviews were conducted until observational and analytical saturation was achieved. Saturation was defined as when no new inductive themes emerged during analysis and a priori or deductive themes were exemplified in the data [33].

Data Analysis
The qualitative data were first coded and analyzed independently by 2 Ctrl Group researchers using a set of foundational structural codes. A structural coding framework, based on the agreed-upon discussion guide from the qualitative interviews, was created for the “top-down” codes; “bottom-up” codes that organically arose from the data were also used. The top-down and bottom-up coding was then refined collaboratively as the researchers progressed through the data set by, for example, splitting preexisting codes into more specific subcodes and combining existing subcodes where appropriate. Once refined, the codes enabled thematic analysis and identification of recurring themes across participants’ accounts and perceptions. Guided by prior research findings [34-36], we adopted thematic analysis to generate an overarching understanding of shared experiences, thoughts, and behaviors across a broad, qualitative data set, to allow both inductive and deductive insights to emerge. This was particularly important because our study aimed to validate preexisting product features (through deductive reasoning) while also allowing respondents to provide suggestions that can inform the development of completely new features (inductive...
reasoning). This type of mixed reasoning would have been very
difficult to achieve if the participants had only been surveyed
about their thoughts around specific product features. The
flexibility allowed by the thematic analyses was also important
because it enabled engagement with both personal accounts of
patients’ experiences and understandings and broad social
constructs (eg, “workflow efficiency” or “medication
adherence”) in slightly different social contexts (ie, different
primary care environments) in the same research [36].

Thematic analysis of interviews was supported by Dedoose
(SocioCultural Research Consultants, LLC), a qualitative data
analysis software tool that enables systematized coding. Audio
recordings from the interviews were uploaded to the web-based
transcription service, Rev. Transcripts were uploaded to Dedoose
[37], the top-down codes created during the coding framework
phase were applied to appropriate excerpts for each transcript,
and new bottom-up codes were created and applied as they arose
over the course of the analysis. High-level analyses were
conducted by reviewing the codes that had been applied to the
interview transcripts. This was done to identify key and
recurring themes from the insight-gathering stage. Interpretive
insights were formulated through subsequent in-depth analyses.
Findings and insights were presented based on code frequency,
salience, and relevance to the research project.

The Standards for Reporting Qualitative Research
recommendations were followed in the reporting of the study
[38].

Ethics Approval
Ethics approval from the AAH institutional review board was
obtained by making amendments to the existing Advocate
Pathway study protocol (approval number: AHC-6680-75000249).

Results

Participant Characteristics
A total of 37 patients living with MDD completed the 18-week primary follow-up period in the pilot feasibility study—19% (n=7) were Black and 38% (n=14) were Hispanic [31]. Overall, 54% (20/37) of patients from the pilot feasibility study (9/18, 50% from the Pathway App arm and 11/19, 58% from the usual care arm) participated in the qualitative interviews in stage 1. In addition, 15 HCPs also participated in the stage-1 interviews. Overall, 53% (8/15) of HCPs in the group were previously involved in the pilot study, 25% (2/8) of whom were assigned to the PathwayApp arm and had used the app to derive patient reports. A total of 36 HCPs from primary care medicine with experience in managing people with depression participated in the stage-2 interviews (Multimedia Appendix 3). These HCPs were sampled for diversity of roles both within and across sites, and it was found that 33% (12/36) were physicians, followed by certified medical assistants (8/36, 22%), registered nurses (6/36, 17%), and licensed practical or advanced practice nurses (6/36, 17%).

Users’ Experiences With and Perceptions of the Pathway App

Overview
Thematic analysis identified major themes to describe the users’ experiences with and perceptions about the Pathway App, which included functionality, support, and patient-provider engagement, along with subthemes such as ease of use, utility, reliability, motivation, reducing burden, communication, understanding, and shared vision (Tables 1 and 2).
Table 1. Experiences with and perceptions of the *Pathway App* among patients living with major depressive disorder and health care providers.

<table>
<thead>
<tr>
<th>Themes, subthemes, and the Pathway App features</th>
<th>Participant quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Functionality</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Ease of use</strong></td>
<td></td>
</tr>
<tr>
<td>Side effects</td>
<td>“It’s really easy to see how your meds are affecting you…” (Patient; app)</td>
</tr>
<tr>
<td>Mood</td>
<td>“It was easy to use because it had good options...it’s hard to pinpoint how you’re feeling on a numbers scale...easier when you’re explaining it.” (Patient; app)</td>
</tr>
<tr>
<td><strong>Utility</strong></td>
<td></td>
</tr>
<tr>
<td>Medication tracking</td>
<td>“...very helpful because if you’re really busy you get notifications that remind you to take it. I used it consistently. I forget a lot and that thing would pop up...everyone needs a reminder especially if you’re on new meds or daily meds.” (Patient; app)</td>
</tr>
<tr>
<td>Side effects</td>
<td>“A lot of mine were on there and it was super easy to use...as time went on I realized I wasn’t having them anymore.” (Patient; app)</td>
</tr>
<tr>
<td>Mood</td>
<td>“It’s great data to see and it’s nice that you can track that, especially with the medication in compliance and the side effects.” (RN)</td>
</tr>
<tr>
<td><strong>Reliability</strong></td>
<td></td>
</tr>
<tr>
<td>Mood</td>
<td>“I think this is beneficial, especially if you see the data and notice, for example, on Saturday the patient seems to be down—what’s happening on this day from a psychological perspective...if they’re always down then maybe you need to increase medication...this gives me objective data.” (HCP; study)</td>
</tr>
<tr>
<td>Medication tracking</td>
<td>“We can know definitively what's going on with the medication and not rely on the patient to give us a history, because they’re more inclined to be honest with their phone on a day-to-day basis and not give us generalities when they come in a month later...This definitely would give us a better picture. More accurate.” (LPN)</td>
</tr>
<tr>
<td><strong>Support</strong></td>
<td></td>
</tr>
<tr>
<td>Motivation</td>
<td></td>
</tr>
<tr>
<td>Medication</td>
<td>“It actually makes you feel like you’re talking to somebody...It lifted my spirits at the end of the day and it’s easy to navigate...very useful because it will make me think...” (Patient; app)</td>
</tr>
<tr>
<td>Side effects</td>
<td>“Sometimes I was looking forward to tracking how I was feeling to keep track of side effects...I really had not been on this type of medication before...so this was just to learn which side effects I was actually having and to tell the difference between the medication to see which would give me least side effects...so it just helped me figure out maybe this is not the best type to be taking...” (Patient; app)</td>
</tr>
<tr>
<td>In-app report</td>
<td>“I like the visual—it helps you remember you need to take your meds every day...it’s very useful to see side by side and broken down by week.” (Patient; usual care)</td>
</tr>
<tr>
<td>Goal setting and tracking</td>
<td>“Especially for young people it’s important to try and remind them. The look is nice...[it has a] pleasant appearance...and provides feedback which is helpful...it looks encouraging too, there’s positive reinforcement...” (HCP; nonstudy)</td>
</tr>
<tr>
<td>Patient education</td>
<td>“I think it gives you something to look forward to. It helps when you’re dealing with depression...gives you a purpose.” (Patient; app)</td>
</tr>
<tr>
<td></td>
<td>“I like to be able to say that I almost got there or I made a little progress...It would help motivate you...” (Patient; usual care)</td>
</tr>
<tr>
<td></td>
<td>“Almost got there’ is good language, it’s supportive...It’s really a form of motivational interviewing and it allows you to go a little further with more information behind you...” (HCP; nonstudy)</td>
</tr>
<tr>
<td><strong>Patient education</strong></td>
<td>“This helps you focus back on yourself. I’m jealous...I want to use it...I was already thinking of this as an addition.” (Patient; app)</td>
</tr>
<tr>
<td></td>
<td>“I love this, it feels very enticing. It would definitely keep me engaged...[it’s] peaceful and motivating.” (Patient; app)</td>
</tr>
</tbody>
</table>
### Themes, subthemes, and the Pathway App features

<table>
<thead>
<tr>
<th>Participant quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reducing burden</strong></td>
</tr>
<tr>
<td><strong>HCP visit preparation</strong></td>
</tr>
<tr>
<td>• “As patients we always write down all our questions, but the doctor only has so much time—a lot of time is just the provider asking a lot of questions, trying to get down to the main thing—this will cut time on the talking...[It’s] easier for shy patients to point out their concerns to their provider. I think doing the questions before the appointment, will actually help” (Patient; app)</td>
</tr>
<tr>
<td>• “I really like that because I’m seeing it almost as a time saver...I’m trying to tease out information and giving them the opportunity to process things ahead of time you’ve already got that foundation...I feel like that would be useful to me and the patient...” (HCP; study)</td>
</tr>
<tr>
<td><strong>Side effects</strong></td>
</tr>
<tr>
<td>• “…This was helpful especially when you go to the doc because you don’t have to try and remember three or four weeks ago...It’s actually nice to be able to tap in side effects without having to go to the doc’s for three hours just to tell them you have dry mouth or fatigue.” (Patient; app)</td>
</tr>
<tr>
<td><strong>Patient education</strong></td>
</tr>
<tr>
<td>• “I really like this because I don’t take the time to do this with the patient. I think patients would engage with it...Aside from treatment the whole behavioral therapy component is key. I think it would make my life a lot easier because it takes a lot of time out of clinical practice.” (HCP; nonstudy)</td>
</tr>
</tbody>
</table>

---
aRN: registered nurse.
bHCP: health care provider.
cLPN: licensed practical nurse.
Table 2. *Pathway App* and its effect on patient-provider engagement.

<table>
<thead>
<tr>
<th>Themes, subthemes, and the Pathway App features</th>
<th>Participant quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Communication</strong></td>
<td></td>
</tr>
<tr>
<td>Goal setting and tracking</td>
<td>“It helps make it more intimate where they’re not just trying to medicate you...it’s outside of just looking at you as a patient...it makes it more personal...” (Patient; app)</td>
</tr>
<tr>
<td></td>
<td>“I love it...this is so exciting for me! This is what I try to do manually but this would be so much more effective...you deepen the conversation.” (HCP; nonstudy)</td>
</tr>
<tr>
<td></td>
<td>“It’s important for the physician to have that information. It helps to hold the patient accountable...to look in black and white and have an honest conversation with the doctor...” (Patient; usual care)</td>
</tr>
<tr>
<td>Side effects</td>
<td>“You can say to someone let’s see what happens after a month. It makes it easier to attribute side effects to meds or other things...which supports the conversation with the patient.” (HCP; nonstudy)</td>
</tr>
<tr>
<td>Patient education</td>
<td>“Going up a hill is a good visual—things seem hard but if you pick a goal and pick away at it...I like the language of getting back to yourself because you’re not yourself and you don’t feel yourself...This would be useful to me...I would want to know the content so I could have conversations and would know where to focus...this would balance the visit out a little bit...I love it.” (HCP; nonstudy)</td>
</tr>
<tr>
<td><strong>Understanding</strong></td>
<td></td>
</tr>
<tr>
<td>For all tracking features</td>
<td>“Great idea...you don’t have to remember every single feeling, every single side effect...if they see an issue during that time they can go right to that point and not waste 35 minutes trying to figure out how you were possibly feeling and trying to remember way back to that day...they just have it right there...and they can say OK, you were feeling this particular way, how are you now?” (Patient; app)</td>
</tr>
<tr>
<td></td>
<td>“I would 100% use this with my patients...[This is] especially good for someone starting or changing meds...now you could know how meds affect sleep, energy...I would want time to digest this so would be good to receive prior to any appointment and to receive on an ongoing basis.” (HCP; nonstudy)</td>
</tr>
<tr>
<td></td>
<td>“There’s a very good indication over time of how someone’s feeling instead of just at this visit when they walk in and [are having] a bad day or a good day.” (RN)</td>
</tr>
<tr>
<td><strong>Shared decision-making</strong></td>
<td></td>
</tr>
<tr>
<td>For all tracking features</td>
<td>“I think it would engage the patient more. They could see that I’m getting the data from them and taking it seriously, and that someone is interested in their condition and is monitoring their condition, so I think it would be good for the patient.” (MD)</td>
</tr>
<tr>
<td></td>
<td>“If I’m seeing it before the patient came in, the PHQ-9 is already done so I don’t have to ask those questions which saves me a little bit of time, and then I would get into ‘looks like you’re still doing not as well as we’d like to,’ so I would go back into shared decision-making.” (MD)</td>
</tr>
</tbody>
</table>

*aHCP: health care provider.*  
*bRN: registered nurse.*  
*cMD: doctor of medicine.*

**Functionality**

Functionality was a major theme that emerged from the interviews, with 3 associated subthemes—ease of use, utility, and reliability. Patients living with MDD and HCPs emphasized ease of use and overall value as 2 of the most beneficial app features. Participants appreciated the simplicity of the design, which made it easy to understand and use the app effectively, and many reported that the app provided clinically meaningful information, such as tracking of symptoms, mood, and medication adherence, which helped them manage their symptoms more effectively. HCPs viewed mood, medication, and side effect tracking to be clinically valuable, owing in part to the continuous and direct input from the patient.

**Support**

Support was an important theme that emerged, which could be split into 2 subthemes—motivation and burden reduction. Patients living with MDD reported that the goal-setting and goal-tracking features provided emotional support and motivation through regular reminders and encouragement, which helped them stay engaged in their treatment. Similarly, HCP visit preparation was noted as very useful by patients living with MDD. Many reported that the ability to prepare for HCP visits using the *Pathway App* reduced their anxiety and improved the ease and accuracy of communication with their HCP.

HCPs felt that the visit preparation feature reduced their administrative burden and improved the efficiency and effectiveness of appointments. Similarly, HCPs stated that patient education features reduced the education burden on them, providing a constructive focus for patients living with MDD and improving patient-provider interactions.

**Patient-Provider Engagement**

Patients living with MDD reported that the app helped to improve their interactions with their HCP by providing
easy-to-use tools for tracking symptoms and progress. They also appreciated the ability to share their data with their HCP, which they said helped to facilitate better communication and collaboration. Patients and HCPs found that the goal-setting and goal-tracking features of the app supported clinical conversations while also helping patients focus on their treatment goals. HCPs reported that patients may remain engaged in their care by knowing that their input and concerns are being taken seriously, thereby promoting SDM.

Both patients living with MDD and HCPs expressed positive feedback about the patient education feature of the Pathway App. They found the feature to be visually appealing and engaging and deemed the content to be suitable in length and depth. Furthermore, HCPs reported that the education feature would facilitate productive conversations with patients.

**Concerns**

Concerns about the Pathway App were that it was difficult to use and had a lack of interactivity, with subthemes such as being confusing, repetitive, and time consuming and lacking workflow integration (Table 3). The initial well-being tracker, a visual analog scale from 0 to 100, received the most critical feedback from patients living with MDD, as many of them found the construct and scale to be unnecessarily complicated. HCPs also expressed skepticism about the clinical utility of the well-being tracker feature and had concerns about the ease with which they could interpret the responses. Both HCPs and patients suggested changes to include a simple response format and visual indicators to support interpretation. The cognition 2-back feature also generated negative reviews, with both patients and HCPs expressing their frustration, as many found it to be confusing and anxiety inducing, and they expressed skepticism about the usefulness and interpretation of data. Participants also expressed frustration about the daily repetition of certain questions and expressed concerns about disengagement.
Table 3. Concerns about and recommendations for the *Pathway App* by patients living with major depressive disorder and health care providers.

<table>
<thead>
<tr>
<th>Concerns</th>
<th>Categories, themes, subthemes, and the <em>Pathway App</em> features</th>
<th>Participant quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cumbersome to use</td>
<td>Pathway report</td>
<td>• “This is helpful but I would really love a 1-page analysis report because I don’t have a lot of time...” (HCP&lt;sup&gt;a&lt;/sup&gt;; nonstudy)</td>
</tr>
<tr>
<td>Time consuming</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confusing</td>
<td>Well-being tracker</td>
<td>• “I didn’t find it as useful as the others...I just left it at 50...I don’t know why. It was just a little confusing.” (Patient; app)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• “I did like it but it’s a little too complicated...what is the point of saying I feel like 86 or 62 today? A 1-10 scale would have been easier to use.” (Patient; app)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• “That’s a very large range. What’s good? What’s bad? Is 80 good? Sometimes giving too large a range is too much for patients to think about. What does that really mean for me?” (HCP; study)</td>
</tr>
<tr>
<td>Cognition 2-back</td>
<td></td>
<td>• “I got aggravated. I didn’t understand the whole concept and I failed. It never explained the purpose. I have no idea what it was trying to do...it was so vague, it just said ‘do you want to start?’ Maybe I felt I failed because of the scores...I didn’t get a lot of greens...I stopped doing them...I didn’t understand why or what I was doing...” (Patient; app)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• “The 2-back thing needs to go or explain it more or more clarification—I wouldn’t take it away completely; it just needs more meaning behind it.” (Patient; app)</td>
</tr>
<tr>
<td>Repetitive and redundant</td>
<td>Side effects</td>
<td>• “...Does it have to be on a daily basis or could it be weekly? I could see it being redundant...” (HCP; study)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• “[I] would caution against patients feeling overwhelmed by the options if they saw them all at the same time.” (HCP; study)</td>
</tr>
<tr>
<td>Mood</td>
<td></td>
<td>• “These questions are really good but every day? I was over it by the third day in a row. If I’m depressed, I don’t want to be reminded...If I’m having a good day, I don’t want to trigger it...[so] I would ignore it...it’s a lot...reminding myself I feel so low all the time...” (Patient; app)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• “...It seems like a lot to ask about well-being and mood...I’m less concerned about daily stuff and more concerned with weekly trends...my intuition is that you don’t need this every day and I would worry about overloading the patient.” (HCP; study)</td>
</tr>
<tr>
<td>Lack of interactivity</td>
<td>In-app report</td>
<td>• “I think you should be able to put in your time frame and ask the app to show me my progress over this amount of time. Make it a fun thing—a metric or a tab where people can see progress over time in a fun way...People are motivated by progress, right?” (Patient; usual care)</td>
</tr>
<tr>
<td>Recommendations</td>
<td>Workflow integration</td>
<td>• “...a 1-page report regardless of the period of time since the last appointment that could be faxed or e-faxed back into the EMR...” (HCP; nonstudy)</td>
</tr>
<tr>
<td></td>
<td>Pathway report</td>
<td>• “I might not be able to look at it before the visit. I can look at it quickly with the patient there as long as it’s simple for me to read, I can scan and we can have a conversation...ideally this would be linked to EMR...integration would help, it would give me more time to talk to them.” (HCP; study)</td>
</tr>
<tr>
<td></td>
<td>HCP visit preparation</td>
<td>• “The sharing it with your physician part makes me raise an eyebrow because that has to be very seamlessly integrated in a way I’m not sure is possible currently...so unless the progress concerns are uploadable into the EMR and integrated into a physician’s workflow...it’s going to be really difficult to follow up on the fact that you told my patient that I have this information...if that [EMR integration] is possible that would be fantastic because it takes care of some of my documentation too because it’s as if I’ve asked these questions even though I haven’t.” (HCP; study)</td>
</tr>
<tr>
<td></td>
<td>Goal setting and tracking</td>
<td>• “If I’m entering anything in my computer, I don’t want to retype, I want it to automatically link. I don’t want to spend any more time on the EMR. I feel like I’m on it 24 hours a day already. If I’m typing in goals, I want it in my progress notes.” (HCP; study)</td>
</tr>
</tbody>
</table>
Participant quotes
Categories, themes, subthemes, and the Pathway App features

### Increase interactivity

**Patient education**
- “Is it going to elaborate more? Like articles and things you could work on or... informational things like why you might be feeling the way that you’re feeling... like help you, make you think a little, try that thing, see if that helps me or even if it was something small like today—have you thought of going outside and taking a deep breath in nature... or like have you sat down and meditated for 5 minutes? It doesn’t have to be a lot, but little things to guide you...” (Patient; app)
- “One thing I haven’t seen is little videos or motivational stories or patient stories or different things to engage patients... or other vendors that offer CBT... or Reddit group support or offers to join support groups...” (HCP; nonstudy)

**Explanation and visualization**

**Cognition 2-back**
- “The 2-back thing needs to go or explain it more or more clarification—I wouldn’t take it away completely; it just needs more meaning behind it.” (Patient; app)

**Well-being tracker**
- “Anything you can make simpler you should make simpler.” (HCP; study)
- “0-100 is a lot of in between... I don’t know how I would rate myself. I just think of the smiling faces [in the hospital]... they ask you on a pain scale of 0-10.” (Patient; usual care)

### Recommendations

For the well-being tracker, participants suggested a simple response format and the inclusion of visual indicators to support interpretation (Table 3). A better explanation and provision of an alternative cognitive exercise were suggested by participants for the cognition 2-back feature. Providing examples of goals and assisting patients living with MDD with setting their own goals independently of their care teams were among the HCP-suggested changes to the goal-setting feature. Both HCPs and patients suggested changes to goal tracking, such as options to set goal reminders, record goal progress, and provide rewards to motivate goal attainment.

Many of the HCP-suggested changes included an emphasis on the value of EHR integration for easy access and to save time. They suggested that features such as HCP visit preparation, goal setting and tracking, and the Pathway report would benefit from integration with the EHR to be more useful and effective for HCPs. In addition to workflow integration, HCP-suggested changes included increasing the interactivity and variety of the given content for patient education, while also making the value and purpose of the educational materials clear.

#### Optimization of the Integration of the Pathway App Into a Primary Care Clinical Workflow

**Overview**

Care team members made several recommendations regarding how to facilitate the integration of the Pathway App into clinical workflows. These included product-specific recommendations and suggested changes to existing workflow scopes. Key themes that emerged were the importance of MBC, SDM, educational needs for the care team, and patient education (Table 4).
Table 4. User perceptions and experiences regarding the integration of the *Pathway App* into a primary care clinical workflow.

<table>
<thead>
<tr>
<th>Categories, themes, and sub-themes</th>
<th>Participant quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Perceptions</strong></td>
<td></td>
</tr>
<tr>
<td>Importance of MBC&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>PROs&lt;sup&gt;b&lt;/sup&gt;</td>
<td>“It definitely gives you a tracking method. You’re able to really see over a period of time how the patient’s doing, which is key, especially for a depressed patient. You may only see them once every 6 months, which again, you don’t get the exact data that you need during that period...you’re really relying that the patient will follow up regularly...If not, at least the app is giving you an idea of what’s going on.” (RN&lt;sup&gt;c&lt;/sup&gt;)</td>
</tr>
<tr>
<td></td>
<td>“It just means making me much more aware of that because I don’t always have the time to dig into this with my patients. I think it makes me a better PA&lt;sup&gt;d&lt;/sup&gt; overall if I’m able to touch on these things. And sometimes it’s just nice to have a reminder right there in front of you while you’re talking to your patients.” (PA)</td>
</tr>
<tr>
<td></td>
<td>“I think it’s nice to actually see it, especially now that you’re tracking it over a period of time and you can see how it’s increasing, decreasing things of that nature. So I think that’s a great method to have it like that. And it’s not too much information, it’s just enough data that shows me where they’re going.” (RN)</td>
</tr>
<tr>
<td></td>
<td>“It’s really just making sure that they’re progressing or their symptoms are improving with their medication. Again, some patients miss visits even if they’re supposed to follow every 2 months and then they don’t make it to that 2-month visit. If Pathway can communicate with that patient and I can see the score, I’ll be able to correlate that with the effectiveness of the medication.” (MD&lt;sup&gt;e&lt;/sup&gt;)</td>
</tr>
<tr>
<td>Shared decision-making</td>
<td></td>
</tr>
<tr>
<td>Goal setting</td>
<td>“You’re able to track if the patients are really adhering to the goals that you set. And it’s more frequent versus me asking them once or twice in the office. This is a more frequent check...” (RN)</td>
</tr>
<tr>
<td></td>
<td>“I think it’s really helpful. It’s going to definitely engage the patient with their care and simplify, not just send them on their way. ‘This is what I want you to do.’ It’s tangible, it’s on your phone, there’s no question as to what was asked or talked about at the visit. It’s right there. I think it will be great for patients.” (LPN&lt;sup&gt;f&lt;/sup&gt;)</td>
</tr>
<tr>
<td>Side effects</td>
<td>“Well, I always ask them what side effects they’re having, but I guess when they come in, I can reinforce they’re side effects, or is it their illness itself based on what I’m seeing here. So...the insomnia, well, I’ll ask them, how’s the medication working, and they might say it’s working well, but they might not say that they’re having insomnia, so I can get that from here. So that would be helpful.” (MD)</td>
</tr>
<tr>
<td></td>
<td>“I would go back into the whole shared decision-making...We tried sertraline before. It looked like you did have some side effects, although they went away. How would you feel about going up on the dose of the medication because I’d like to try to tap out before I switch?” (MD)</td>
</tr>
<tr>
<td><strong>Concerns</strong></td>
<td></td>
</tr>
<tr>
<td>Overwhelming</td>
<td>“As a physician I worry...with this information, what if I miss something? Because there’s so much information; if I don’t know where to find it, how to use it...especially where to find it...In a timely, effective, efficient way, I’m afraid that, what if I miss something that is crucial for this patient?” (MD)</td>
</tr>
<tr>
<td>Educational support</td>
<td></td>
</tr>
<tr>
<td>Cognition tests</td>
<td>“Maybe there was some training for the provider so it’s like actually this tracks really well with how their cognitive performance is doing with depression. Then [I’d say] ‘Okay, you got me. I’m in.’” (MD)</td>
</tr>
<tr>
<td>PROs</td>
<td>“I don’t know enough about this WHO&lt;sup&gt;g&lt;/sup&gt; or the PDQ-D-5&lt;sup&gt;h&lt;/sup&gt;, what these numbers mean. And if they’re getting better, or worse. I’m just not familiar with those. And then the same thing with these 2. I guess I just don’t know enough about these 2 scores.” (APN&lt;sup&gt;i&lt;/sup&gt;)</td>
</tr>
<tr>
<td><strong>Recommendations</strong></td>
<td></td>
</tr>
<tr>
<td>Patient education</td>
<td></td>
</tr>
<tr>
<td>Educational resources for patients</td>
<td>“I’ll give them handouts and such that I find on say, UpToDate, or embedded in our EMR&lt;sup&gt;j&lt;/sup&gt;.” (PA)</td>
</tr>
<tr>
<td></td>
<td>“I think we’re always weary, but I think more knowledge is better generally with patients, rather than them being surprised. [...] I would say drugs, actions, and side effects would be useful.” (MD)</td>
</tr>
</tbody>
</table>

---

<sup>a</sup>MBC: measurement-based care.

<sup>b</sup>PRO: patient-reported outcome.

<sup>c</sup>RN: registered nurse.

<sup>d</sup>PA: physician assistant.
Importance of MBC

Although many primary care professionals within Advocate Aurora primary care perceived MBC, collaborative care, and SDM as important, various barriers limited their inclusion in day-to-day clinical workflows. Care teams agreed that Pathway administration of the PHQ-9 builds on and improves current clinical practices. HCPs viewed the PHQ-9 as a key component of current clinical practice; therefore, its continuous use to track patients’ treatment progress outside visits, as enabled by Pathway Platform, was perceived as clinically useful.

Shared Decision-Making

Tracking collaboratively devised goals was perceived as useful for the care team, with many also stating that this would help engage people living with MDD in their own care. HCPs reported that the presentation of patient-reported outcome (PRO) trajectories for medication adherence, PHQ-9 scores, and side effects provided a clinically useful view of how each patient’s condition has changed over time, which, in turn, supports better clinical decision-making. HCPs suggested including more information about patient care in the Pathway-EHR interface, such as medication refill data for comparison with reported adherence.

Workflow Integration and Data Visibility

Interviews also highlighted the importance of understanding care team needs, such as interpreting PRO measures and trend visualizations, to ensure that Pathway Platform can support care team workflows. Data about medication adherence, PHQ-9 scores, and side effects were perceived to be the most clinically important PROs. However, there were also concerns that Pathway data would be overwhelming and may lead to key data being missed.

Educational Needs

Many expressed low familiarity and desire for education regarding PROs other than PHQ-9 and clinical use of cognitive tests. In addition to their own education, the care team members also expressed a desire for patient-directed educational materials. Many care team members spoke about patients living with MDD being provided with educational handouts at their respective practices (including handouts printed from web-based searches, those that have been externally printed, and those taken from web-based medical reference platforms and the EHR). Many care team members stated that, from the list of educational topics (which included understanding one’s diagnosis, treatment options, side effect management, goal setting, and patient engagement), they were most interested in accessing materials about helping patients understand their diagnosis and their available treatment options. Care team members also thought that patients would benefit from educational resources about the types of local behavioral health resources that are available to them and information about what type of role they can play in their depression care and how often they should follow-up with their physician.

Discussion

Principal Findings

Qualitative research and thematic analyses conducted in this study allowed us to capture the user experiences and perceptions of patients living with MDD and their care teams. These results can guide researchers and app developers in designing effective digital health interventions that will be readily accepted by their intended end users [30].

Most of the currently available apps developed for depression management have only been assessed for effectiveness in a research setting and have not been integrated within clinical workflows; this has resulted in a lack of adoption by care teams and broad health care systems [26]. Incorporating the voices of people living with MDD and their care teams into the product development process aligns with the broad paradigm shift toward patient-focused decision-making and SDM between patients and their providers, ultimately leading to high-quality, fully informed, and preference-based treatment plans [39]. Our analyses revealed that both patients and their care teams perceived the remote tracking of PROs via digital tools to be clinically useful and reliable. Other highlights of our study included the need to enhance the flow of real-time data shared with the care team and the need to integrate within the care team workflow, including real-time sharing of the patient’s app data within the EHR. Results from our analyses also highlighted the need for care team education about MBC and SDM and about how to use the Pathway App to improve these processes by using features such as visit preparation and collaborative goal setting and tracking. Using the broad insights gathered from our thematic analyses, we were able to understand, iterate, and integrate a digitally enabled platform, Pathway Platform, into a primary care setting in the United States. The first iteration of the Pathway App included PRO measures related to depression, well-being, cognitive symptom tracking, medication adherence, and side effects [31]. Pilot results confirmed the feasibility of using the Pathway App among patients living with MDD and showed a trend in enhanced patient activation in the app arm, albeit in a small sample size [31]. Building on the results from the pilot study, Pathway Platform was developed to consist of 3 components: the latest iteration of the Pathway App; EHR-integrated, real-time, patient-level data sharing; and educational programming that is both care team and patient facing (including a web-based educational resource center that describes the utility of Pathway Platform to the care team through reading materials, presentations, and videos; Multimedia Appendix 4). The current version of the app prompts patients
living with MDD to complete the following scales every 2 weeks: PHQ-9 [40] and Perceived Deficits Questionnaire—Depression [41] to assess depression status and subjective cognitive impairment, World Health Organization Well-being Index [42] to assess quality of life and emotional well-being, and Digit Symbol Substitution Test [43] as an objective measure of cognition to assess working memory and processing speed.

The Pathway App also includes a daily “evening check-in” to collect information about medication adherence and side effects. Data collected by the Pathway App are electronically transmitted and stored in an EHR-integrated web interface. These data are accessible to the care team and provide a longitudinal summary that may assist them in clinical decision-making and overarching depression management. Care team members can view these data either before or during the clinical visit and then use the data to collaboratively discuss future treatment decisions with the people they are treating. A web-based educational training program for primary care team members was also developed by using evidence-based medicine, building on the concepts of MBC and SDM, as they relate to depression management.

In addition, audit and feedback sessions will be conducted to benchmark performance measures, reflect on current clinical practice and improvement strategies, and set team-based action plans. Specific training sessions were conducted for the care team members about how to instruct patients to use Pathway Platform and how to use EHRs to view data collected via Pathway Platform. A training manual was also developed for patients living with MDD that describes the functionality of Pathway Platform and how to interpret their data.

Pathway Platform was cocreated with input from all users (people living with MDD, care teams, health system information technology personnel, and study collaborators), along with input from software developers and health-technology product development researchers, to optimize usability, utility, iteration speed, and integrated system performance and to ultimately enable nuanced care focused on SDM and MBC (Multimedia Appendix 5). Continued reassessment based on user feedback has allowed for fast iterations, optimized system performance, and sustainability [39]. This user-centric approach has, in turn, led to an enhanced digital platform to improve treatment outcomes by supporting an expanded understanding of MDD treatment, bolstering care team workflows, and providing patients with additional support throughout their treatment journey.

Limitations and Future Directions

A limitation of our study is its small sample size of patients with MDD and their care teams. In addition, among the care teams, only 33% (12/36) were clinicians. Future studies would benefit from an even split between clinicians and nurses in the sample. Moreover, this was a single-provider network study, and the results of this study may have limited generalizability. In addition, although the design and development of Pathway Platform were guided by a deep understanding of care team workflows, the extent to which clinical workflows can be modified to accommodate the adoption of Pathway Platform will ultimately depend on the clinical team.

Furthermore, although thematic analyses offered the necessary tools for organizing, interpreting, and transforming data without the need for separate theories, the depth of our conclusions may have been increased by additional methods such as modeling and theory building [36]. However, creating an overarching or generalizable theory to explain the way people reacted to the product’s feature set as a whole [44] would have gone beyond the primary interests of the study, which were focused on understanding and analyzing specific feedback such that it could be mindfully applied to iterations of the product features. We also used a highly systematized and enumerative approach to coding and generating themes, as recommended by the guidelines [34]. In addition, we minimized interpretive inconsistencies by using a single code tree and 2 analysts to evaluate each other’s work for analytical consistency. The improved iteration of Pathway Platform is being evaluated in an ongoing large-scale implementation study (Use of a Digitally Enabled App With Clinical Team Interface in the Management of Depression; NCT04891224). The study will include up to 200 patients at 20 primary care clinics. The implementation study aims to test the scaling and integration of Pathway Platform, along with educational interventions, at multiple primary care sites within the AAH system, with the primary objective of determining improvement in adherence to MBC practices [32]. Results are expected to provide insights into the improvements in clinical workflows that are necessary to enhance collaborative care, depression management, clinician and patient experience, adherence to medication, patient-provider engagement, and depression outcomes in the primary care setting [32]. EHR integration and how it enables decision-making, and efficiencies with current AAH information technology platforms, such as ease of access of data in real time by the care team, will also be assessed. Together, insights from this study will allow further amendment of workflows to ensure the optimal use of Pathway Platform [32].

Providing effective care for MDD has become more important than ever, with the prevalence of symptoms of anxiety disorder and depressive disorder having increased more than 3-fold in the United States during the COVID-19 pandemic [45,46]. Furthermore, people with few social and economic resources had high likelihood of exhibiting depression symptoms during this time [46]. Digital tools are therefore increasingly relevant in the era of COVID-19, owing to increased use of telehealth services to facilitate access to care [47]. In addition, low-cost interventions, such as digital tools, could provide increased monitoring and improved services to at-risk populations [15,48].

Moreover, future research methodologies, analysis protocols, and publications should provide a more explicit account of the impact of people’s social intersectionality on their perceptions about remote MDD monitoring. The imperative for this is heightened by the increasing attention that reimbursement entities are paying to the way any given intervention can help minimize the negative effects that social determinants have on treatment outcomes [49].

Conclusions

Ongoing communication with patients with MDD and their care teams (cocreation) provided an opportunity for continued
reassessment and developmental iteration of *Pathway Platform*. These insights included the need for rapid communication of updated and current patient data with the care team, integration of the app into the MDD care pathway via the EHR, and education of the care team about the interpretation and use of these data. This cocreation model using qualitative research findings has resulted in fast iterations and optimized system performance and will allow for eventual sustainability outside the research environment. Future development of *Pathway Platform* will continue, consistent with the evolving needs of people living with MDD and their care teams.

**Acknowledgments**

This study was funded by Takeda Pharmaceuticals U.S.A., Inc., and Lundbeck LLC. Under the direction of the authors, medical writing assistance was provided by Sangeetha De, PhD, on behalf of Syneos Health Medical Communications LLC. Takeda Pharmaceuticals U.S.A., Inc., and H. Lundbeck A/S provided funding to Syneos Health for support in writing this manuscript.

**Data Availability**

The data sets generated and analyzed during this study are available from the corresponding author upon reasonable request.

**Authors' Contributions**

MM, RK, CK, CB, BF, JK, LC, and DEK contributed to the conceptualization of the study. MM, CB, BF, JK, LC, and DEK supported the development of the study methodology. BF and JK supported the software components of the study and performed data validation. RK, BF, JK, AS, and LZ performed the formal data analysis. BF, JK, and DEK contributed to data collection and provision of study resources. RK, BF, JK, AS, and LZ contributed to data interpretation. MM, RK, CB, BF, JK, LZ, and AR contributed to data visualization. MM, RK, CK, CB, BF, JK, and DEK provided oversight for research planning and execution. MM, RK, CK, CB, BF, JK, and AR managed, coordinated, and executed research activities. All authors contributed to draft review and critical revision of the manuscript and approved the final version to be published.

**Conflicts of Interest**

MM is an employee of Takeda Pharmaceuticals U.S.A., Inc., and receives annual stock options. LC and AR were employees of Takeda Pharmaceuticals U.S.A., Inc., at the time of the study. CB was an employee of Advocate Aurora Health at the time of the study and is currently an employee of Takeda Pharmaceuticals U.S.A., Inc. RK, CK, and DEK are employees of Advocate Aurora Health. DEK has received remuneration from Takeda Pharmaceuticals U.S.A., Inc., for activities unrelated to the conduct of the study. JK and BF are employees of Ctrl Group. JK is also a founder and owner of Fora Health, the software used in this study. BF is also a director of Cognition Kit. LZ and AS were employees of Ctrl Group at the time of the study.

Multimedia Appendix 1

The treatment journey for patients with major depressive disorder (MDD).
[DOCX File , 534 KB - humanfactors_v10i1e42768_app1.docx ]

Multimedia Appendix 2

Screenshot of the *Pathway app* and a sample report from the pilot study.
[DOCX File , 1043 KB - humanfactors_v10i1e42768_app2.docx ]

Multimedia Appendix 3

Details of health care provider sample in stage-2 interviews.
[DOCX File , 50 KB - humanfactors_v10i1e42768_app3.docx ]

Multimedia Appendix 4

Design of *Pathway Platform*.
[DOCX File , 156 KB - humanfactors_v10i1e42768_app4.docx ]

Multimedia Appendix 5

User-centered design approach.
[DOCX File , 68 KB - humanfactors_v10i1e42768_app5.docx ]

**References**

1. Depression. World Health Organization. 2021 Sep 13. URL: [https://www.who.int/news-room/fact-sheets/detail/depression](https://www.who.int/news-room/fact-sheets/detail/depression) [accessed 2022-06-01]


49. Solomon LS, Kanter MH. Health care steps up to social determinants of health: current context. Perm J 2018 Dec 22;22(4S) [FREE Full text] [doi: 10.7812/tpp/18-139]

Abbreviations

AAH: Advocate Aurora Health
EHR: electronic health record
HCP: health care provider
MBC: measurement-based care
MDD: major depressive disorder
PHQ-9: 9-item Patient Health Questionnaire
PRO: patient-reported outcome
SDM: shared decision-making

©Maggie McCue, Rasha Khatib, Christopher Kabir, Chris Blair, Ben Fehnert, James King, Alexander Spalding, Lara Zaki, Lambros Chrones, Anit Roy, David E Kemp. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 26.07.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Designing Virtual Natural Environments for Older Adults: Think-Aloud Study

Rikard Lundstedt, MA; Johanna Persson, MSc, PhD; Carita Håkansson, MSc, PhD; Susanne Frennert, MSc, PhD; Mattias Wallergård, MSc, PhD

1Ergonomics & Aerosol Technology, Department of Design Sciences, Lund University, Lund, Sweden
2Division of Occupational and Environmental Medicine, Lund University, Lund, Sweden

Corresponding Author:
Rikard Lundstedt, MA
Ergonomics & Aerosol Technology
Department of Design Sciences
Lund University
Box 118
Lund, 221 00
Sweden
Phone: 46 46 222 79 16
Email: rikard.lundstedt@design.lth.se

Abstract

Background: Spending time in natural environments is beneficial for human health, but many older adults have limited or no access to natural environments. Virtual reality technology may be a means to facilitate nature experiences, and so, there is a need for knowledge on how to design virtual restorative natural environments for older adults.

Objective: The aim of this study was to identify, implement, and test older adults’ preferences and ideas regarding virtual natural environments.

Methods: A total of 14 older adults (mean age 75, SD 5.9 years) participated in an iterative process to design such an environment. We used think-aloud protocols and qualitative content analysis and established questionnaires that targeted usability, affective aspects, and side effects. These data guided the design decisions for incremental implementations of a prototype.

Results: The participants’ preferences included trueness to reality in terms of rendition and behavior; traces of human activity and natural processes that trigger the imagination and provide believability; the ability to roam, explore, and interact with the environment; and a familiar, relatable environment that evokes memories. The iterative design process resulted in a prototype featuring many of the participants’ ideas and preferences, including a seated locomotion technique, animals, a boat ride, the discovery of a boat wreck, and apple picking. The questionnaire results indicated high perceived usability, interest, and enjoyment; low pressure and tension; moderate value and usefulness; and negligible side effects.

Conclusions: We suggested 3 principles for virtual natural environments for older adults: realness, interactivity, and relatedness. Virtual natural environments should also provide a diversity of content and activities to accommodate the heterogeneity in older adults’ preferences. These results can contribute to a framework for designing virtual natural environments for older adults. However, these findings need to be tested and potentially revised in future studies.

(Keywords: virtual natural environments; user-centered design; qualitative method; real-time 3D graphics; older adults)

Introduction

Background

Currently, there is much evidence that spending time in natural environments can be beneficial for cognitive function, mental health, and well-being [1-6]. Unfortunately, many older adults, who may be in particular need of such health benefits, have limited or no access to natural environments. For example, older adults living in residential care facilities may have a diminished ability to go outside because of limitations in functioning that are associated with old age. A few studies have suggested that...
replacements for real nature experiences such as indoor gardens can have positive effects [7, 8]. A study found significant improvements in sleep, agitation, and cognition among 23 institutionalized patients with dementia who were allowed to cultivate and care for easy-to-grow edible plants indoors for 28 days [9]. Virtual reality (VR) has also been suggested as a way for older adults who cannot go outside to spend time in virtual natural environments (VNEs) [10, 11].

A systematic review of indoor nature interventions (indoor gardens, plants, photographs, films, and 1 nonimmersive VR forest presented on a single large screen) for older adults in residential care settings found mixed results [12]. There was not sufficient evidence to recommend such interventions over other interventions or activities. In addition, interventions that involved photos, films, or nonimmersive VR were less effective than interactions with real forms of nature such as indoor gardening. However, immersive VR technology such as head-mounted displays (HMDs), where the viewer is completely surrounded by the virtual environment, has a higher potential to provide a sense of presence (ie, the perceptual illusion of being there) than nonimmersive media such as photos and single-screen displays [13-15]. Thus, although nonimmersive media would be unlikely to override the residents’ sense of indoor presence, a VNE experienced through immersive VR technology may provide residents with the perceptual illusion of being outdoors in nature.

A related study [16] compared different types of mediated nature experiences by depicting tropical coral reef scenes. The results showed that real-time 3D computer graphics via an HMD elicited a greater sense of presence, nature connectedness, and positive affect than filmed footage via a single screen. Another study comparing different delivery methods of psychotherapeutic interventions found reports of greater positive affect, satisfaction, and perceived credibility by participants exposed via an HMD than by those exposed via a single-screen display or using mental imagery [17]. Immersive VR has also been shown to induce physiological reactions in test participants similar to those induced by a corresponding real situation [18]. With this in mind, one can easily imagine the potential of VR to provide immersive experiences of simulated natural environments to care facility residents who have limited or no access to real natural environments. Recently, a few studies were carried out with older adults experiencing VNEs through immersive technology (HMDs). These studies reported some positive results, such as displayed enjoyment and relaxation [19] and positive responses, and that VNEs soothed and evoked memories [20] and improved positive affect and nature connectedness [21].

VNE studies often cite theories of restoration in natural environments [22], such as attention restoration theory [23], stress reduction theory [24], and the biophilia hypothesis [25]. Restoration theories can provide guidance for the design of restorative environments, for example, the 4 components by Kaplan [23]: being away, extent, soft fascination, and compatibility. Nukarinen et al [26] presented a framework connecting restoration theories and the measurement of health outcomes in VNE studies.

However, VR presents both limitations and possibilities that are different from those of a real natural environment. For example, VR cannot yet mimic the complexity, dynamic behavior, and immense detail of a real natural environment. On the other hand, it provides more or less complete control over the form and function of the virtual environment. Hence, a designer of VNEs is faced with choices that are not applicable in a real natural environment. Moreover, a user’s perception of an artifact is colored by the context in which it resides [27]. As VR provides a perceptual but not cognitive illusion [13]—that is, it feels real, but the user is aware that it is not—artifacts in VR are perceived differently from their corresponding objects in actual reality. As an example, a viewer may be impressed by how real a moss-covered rock looks in VR but may think nothing of a similar rock in the real world. Therefore, the knowledge of real restorative natural environments may not be applicable to virtual ones. To our knowledge, there are no frameworks, models, or guidelines for designing VNEs that are based on knowledge generated through a bottom-up approach in a VR context.

In general, older adults are seldom involved in participatory design processes; in particular, they are not involved in the development of new technologies such as VNEs [28]. As a result, older adults are rarely given a voice in the development of technological solutions aimed at them, which in turn may lead to the rejection or nonadoption of the technological solution in question [29].

Research Questions

In this study, we aimed to identify, implement, and test preferences and ideas for VNEs by involving older adults in an iterative design process of a prototype. Thus, we explored the following research questions: (1) What preferences and ideas do older adults have for a VNE? and (2) How can we realize them?

We present our design process, the outcomes of 3 iterations, and some suggestions for what ought to be considered by designers when designing VNEs for older adults. We hope that our description of the explorative design process, along with interpretations and reflections on older adults’ preferences for VNEs, are a valuable contribution for designers and researchers alike and that our study may serve to indicate future directions in the development of VNEs for older adults.

**Methods**

**Overview**

This study was conducted as a user-centered design process of a VNE to elicit the needs and preferences of users, challenge assumptions, and explore design ideas [30]. Users representing the target group were invited to iteratively test and provide feedback on the VNE prototype in a laboratory environment. Data were collected through a think-aloud protocol and questionnaires. Qualitative data were analyzed using an inductive qualitative content analysis method inspired by Graneheim and Lundman [31]. A total of 3 iterations were performed. The results of the data analyses of the previous iterations served to inform the design choices in the subsequent
implementation phases. By iteratively developing the prototype and testing and analyzing the participants' reactions (Figure 1), we were able to generate and test design concepts in small increments and, thus, build the VNE from the bottom up based on user input throughout the process. After completing the final iteration, we continued the analysis with a focus on the underlying threads of meaning running through all the data [31].

**Figure 1.** Overview of the iterative design process.

- **Ethics Approval**
  Ethics approval for this study was obtained from the Regional Ethical Review Board, Lund, Sweden (2017/118).

- **Test Setup**
  The tests were performed in a laboratory environment equipped with recording equipment for video and audio. The VNE prototype was developed and tested on a Windows PC with a GeForce GTX 1080 graphics card and an HTC Vive VR headset. We used the HTC Vive’s room-scale tracking capabilities and dedicated a play area of approximately 3 × 4 m within the laboratory.

  Each test session comprised the following six steps (steps 1-3 were only included in the participants’ first sessions): (1) the participant was welcomed to the laboratory and given an explanation of the background of the study; (2) the participant was invited to undergo a short demonstration session to familiarize themselves with using the VR system; (3) the participant was asked to fill out a background questionnaire covering their previous experiences with VR and natural environments; (4) the participant underwent a concurrent think-aloud test session (described in the Think-Aloud Protocol section) while using the prototype (this session was video recorded); (5) the participant was asked to fill out questionnaires measuring usability, affective aspects, and side effects; and (6) the participant underwent a retrospective think-aloud session (described in the Think-Aloud Protocol section) while watching the recorded material from the concurrent think-aloud session (this session was also video recorded).

- **Participants**
  A total of 14 participants (n=9, 64% women and n=5, 36% men) were recruited via retiree organizations in southern Sweden. Participants’ ages ranged from 69 to 90 years. The mean age was 75 (SD 5.9) years, and the median age was 73 years. The inclusion criteria were being aged >65 years, being able to speak and understand Swedish, having adequate eyesight to watch television, and being able to transport themselves to the laboratory. Exclusion criteria were any propensity for dizziness or motion sickness, dementia, reduced cognitive function, and problems with balance. The sampling was purposeful as the participants were intentionally selected to elucidate the VNE prototype from the perspective of an older person, and the sampling was convenient in the sense that the participants themselves chose to sign up for the study rather than being randomly selected [32].

  In total, 57% (8/14) of the participants took part in the first iteration (Table 1). All of these participants (8/8, 100%) also took part in the second iteration along with 3 new individuals, resulting in 79% (11/14) of the total participants testing the second version of the VNE prototype. In the third iteration, 93% (13/14) of the total participants took part, whereof 23% (3/13) were new recruits and 23% (3/13) had participated in the second iteration.
Table 1. Overview of participants for each iteration in the design process.

<table>
<thead>
<tr>
<th>Iteration</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
<th>P7</th>
<th>P8</th>
<th>P9</th>
<th>P10</th>
<th>P11</th>
<th>P12</th>
<th>P13</th>
<th>P14</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>2</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>3</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

**Think-Aloud Protocol**

During the test sessions, we used an adapted concurrent and retrospective think-aloud protocol [33]. Before each concurrent think-aloud session, the participant was asked to speak freely about their experience while using the VNE and share their ideas and suggestions for how they thought it should be changed or further developed. When necessary, the first author would use directive probing techniques [34] in an attempt to cover aspects of the VNE that the participant had not yet addressed, for example, by asking, “What is your perception of the water?” To allow the participant to go into greater depth in their reasoning, the first author would sometimes attempt nondirective probing techniques [34]. The retrospective think-aloud session was conducted in the same manner, only instead of using the VNE, the participant watched a video recording of the concurrent think-aloud session with the first author (Figure 2).

**Questionnaires**

Although this was primarily a qualitative study, standardized questionnaires (Multimedia Appendix 1) were included in each iteration as a complement to gain an understanding of how our implementations affected the usability, affective aspects, and side effects of the prototype. Usability was measured using the System Usability Scale (SUS) [35], which is a widely used method for measuring the perceived usability of a system. It consists of 10 statements rated by the user on a 5-point Likert scale about a system’s characteristics, including complexity, ease of use, consistency, and learnability, for example, “I found the various functions in this system were well integrated.” Bangor et al [36] created the following adjective rating scale to interpret SUS scores: Worst imaginable, Awful, Poor, OK, Good, Excellent, and Best imaginable. Affective aspects were measured using the Intrinsic Motivation Inventory (IMI) [37], which is a multidimensional measure that can be used for assessing a person’s experience of an activity. The IMI can be customized to only include the dimensions relevant to a particular use. We chose to include the following dimensions: interest and enjoyment, value and usefulness, and felt pressure and tension. We chose to exclude the following dimensions: perceived competence, effort, and perceived choice. Side effects were measured using the Virtual Reality Symptom Questionnaire [38]. The questionnaire consists of 13 questions measuring physical symptoms that may be experienced when using a VR system, such as headache or nausea, rated by the user on a 7-point scale. In their first test sessions, the participants were also asked to fill out a background questionnaire regarding their past and present experiences of natural environments, their previous level of experience with VR, and their state of well-being at the moment.
Analysis

After each iteration, the notes from the think-aloud sessions were analyzed by one of the authors (RL) using the recordings as a reference when necessary. Each idea, preference, opinion, suggestion, or other thought expressed by the participants was coded using an inductive content analysis method inspired by Graneheim and Lundman [31]. By searching for patterns within the codes, categories emerged that guided the decision-making when implementing new functions, content, and other changes in the VNE prototype.

In the implementation stages of each iteration, authors RL and MW discussed the categories (hereinafter referred to as preferences) to reach an agreement on how they should be interpreted from a design perspective, that is, how the preferences could be realized as changes to the prototype. RL then implemented the agreed-upon changes. As the participants reacted to the new features in the subsequent tests, the preferences that had been realized within the prototype were to some extent validated by the participants themselves. They would sometimes build on and develop them through more specific requests, reflections, suggestions, or ideas. Thus, there was a progression and deepening of ideas over the course of the iterations. We attempted to reflect this in this paper by presenting the preferences in the order in which they were realized.

After completing the final iteration, we continued the analysis assuming a bird’s-eye view of the data. By reflecting on the patterns and interrelationships of the preferences, we arrived at a set of principles for VNEs for older adults. All the authors reached a consensus on the principles by means of repeated discussion.

The Initial Prototype

In the construction of the VNE prototype, we used computer-generated real-time 3D graphics via the Unreal Engine (version 4; Epic Games, Inc) game engine. Thus, it generated visuals, audio, and other potential sensory outputs in the moment based on user input and programmed behavior. The virtual environment was a compound of authored and sampled (from the real world) materials such as 3D models, images, and sound files. To leave room for the participants’ ideas, we designed the initial prototype to be a rather bare and simple starting point. As the subject of this study was the limited access to natural environments in one’s own region, the features included in the scene were of the type that one might expect to see in the geographical region in which the study was set. After putting on the HMD, the participants would find themselves next to a lake near a forest (Figure 3) in a scene comprising various 3D models of grass, moss, water, rocks, sand, dirt, trees, leaves, flowers, and natural debris such as old twigs and logs. Other than the water and grass, which were animated to simulate movement caused by the wind, all objects were static. The environment was lit like a sunny summer day, and the sounds of small birds could be heard. The participants entered the simulation standing up and were able to move around freely over the play area using their own bodies as they would in the real world. This freedom to move allowed the participants to turn around and obtain a complete 360° view of the scene, inspect details such as flowers on the ground by moving closer to them, and look behind objects such as rocks and trees.

Figure 3. Screenshots from the initial prototype: (A) detail of the ground, (B) view facing the lake, (C) view of the play area from the lake, and (D) detail of the water.
Results

Overview
The average concurrent and retrospective think-aloud sessions lasted 29 and 23 minutes, respectively. Analyses of the sessions yielded several categories reflecting various topics and containing both the participants’ opinions of the current iteration of the VNE along with ideas, preferences, and suggestions for future versions. These were not necessarily exclusive to a particular iteration; however, we present them in an order that shows the progression of the design process. Hence, for each iteration, we present the ideas, preferences, and suggestions that inspired the implementations made in that particular iteration or that are linked to it in other ways. By doing so, we aimed to show how participation guided the design rather than being tokenistic.

Participants
The participants generally had very little experience with VR but much access to and experience of spending time in natural environments (Table 2). All except 1 of the 14 participants (13/14, 93%) answered “yes” to the question “Do you have access to a garden in connection with your home?” Generally, the participants reported high perceived well-being at the moment. Some background questions were left out of this table.

Table 2. Questions and scores from the participant background questionnaires.

<table>
<thead>
<tr>
<th>Question</th>
<th>Mean score (SD)</th>
<th>Median score</th>
</tr>
</thead>
<tbody>
<tr>
<td>How much experience do you have with using VR(^a) (1=none; 5=very much)</td>
<td>1.1 (0.4)</td>
<td>1</td>
</tr>
<tr>
<td>How much experience do you have of natural environments? (1=none; 5=very much)</td>
<td>3.9 (0.9)</td>
<td>4</td>
</tr>
<tr>
<td>How great are your possibilities to spend time in natural environments? (1=none; 5=very great)</td>
<td>3.9 (1.1)</td>
<td>4</td>
</tr>
<tr>
<td>How often do you spend time in natural environments? (1=never; 5=very often)</td>
<td>3.2 (0.8)</td>
<td>3</td>
</tr>
<tr>
<td>How do you feel? (1=very bad; 5=perfectly good)</td>
<td>4.3 (0.8)</td>
<td>4.5</td>
</tr>
</tbody>
</table>

\(^a\)VR: virtual reality.

The First Iteration
During the first iteration, the participants tested the initial prototype, spoke their preferences and ideas during the think-aloud sessions, and filled out the questionnaires. Figure 4 shows the results of the IMI (Figure 4A) and SUS (Figure 4B) questionnaires. The IMI scores indicated that most participants (7/8, 88%) found the experience to be highly interesting and enjoyable and somewhat useful and valuable to them and did not feel much pressure or tension. The median SUS score was 85, which is just below excellent on the adjective scale by Bangor et al [36]. An analysis of the think-aloud data is presented in the following section.

Figure 4. (A) Intrinsic Motivation Inventory and (B) System Usability Scale results for the initial prototype in the first iteration.

The Second Iteration
To begin the second iteration, we analyzed the think-aloud data from the first iteration. Table 3 presents the participants’ preferences along with suggestions and illustrative quotes from the participants. As can be seen, they preferred more movement, life, and change and the ability to roam and explore the environment but also sit and relax, and they emphasized the importance of authenticity and realism.
To continue with the next step in the second iteration, we proceeded to interpret the participants’ preferences and suggestions in terms of how we could realize them as changes to the prototype. Table 4 presents the implementations along with the preferences they addressed and the reactions of the participants. We added various animals, movement of the trees, the possibility to roam and explore the environment, and a jetty on the lake. As can be seen, we implemented 2 different ways to enable the participants to roam and explore the environment. Although teleportation is commonplace in many VR applications, the experience of teleporting is fundamentally different from walking as one does not perceive any gradual movement, and participants had specifically expressed a wish to take walks. Therefore, we realized the need for a technique closer to walking. An obvious solution was to simply implement forward propulsion in the direction of the forward vector of either the HMD or the handheld controller at the push of a button on the controller. However, many older people are wheelchair users or experience reduced postural stability and may lose their balance and fall because of vection. We speculated that there would be very few residents at care facilities who could manage immersive VR while standing up. We also considered that a seated locomotion technique might accommodate to some degree the participants’ preference for sitting, passive enjoyment, and relaxation.

In light of this reasoning, we implemented a system in which a user could sit down while driving around in the virtual world. This was accomplished by fastening one of the handheld controllers to the back of a swivel chair, allowing the chair’s position and orientation to be tracked by the VR application (Figure 5). In the virtual environment, the chair was represented by a simple 3D model. The virtual chair’s position and orientation were updated in real time to correspond to those of the real chair. To control the throttle, the user could press a button on the other handheld controller while sitting in the swivel chair, which would result in them experiencing forward propulsion in the direction of the chair in the virtual environment. To steer, the user would simply turn the chair in the direction in which they wanted to go using their feet. Thus, the user would not experience circular vection while steering, something that is associated with motion sickness [39].

At the end of the second iteration, the participants tested the new prototype while thinking aloud, reacting to the new implementations and providing new preferences and suggestions. As before, they filled out the questionnaires. Table 4 presents some quotes that illustrate the participants’ reactions. Figure 6 shows the IMI (Figure 6A) and SUS (Figure 6B) scores. The IMI scores indicated that the participants found the experience of the new VNE to be highly enjoyable and interesting, that they did not feel too pressured or tense, and that the value and usefulness of the experience to them was neutral to high. The median SUS score was just above good on the adjective scale by Bangor et al [36]. This indicated that, although the complexity of use increased with the implementation of teleportation and the swivel-chair vehicle, usability remained satisfactory.

### Table 3. Participants’ preferences and suggestions from the think-aloud session in the first iteration.

<table>
<thead>
<tr>
<th>Preferences</th>
<th>Description</th>
<th>Participants’ suggestions</th>
<th>Illustrative quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Movement, life, and change</td>
<td>The environment was perceived as static and sterile and lacked life, change, and movement.</td>
<td>Add animals, changes in the light from cloud movements, and movement in the trees from the wind.</td>
<td>“The nature looks like a still [photograph].” [P4]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“In [real] nature, there is some movement in some way all the time...you see birds, but you don’t see any birds here.” [P5]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“I think it would be positive with something that showed up. There could come a hedgehog, and butterflies, and a bird that flew.” [P3]</td>
</tr>
<tr>
<td>Roaming and exploring</td>
<td>Participants wanted to move beyond the play area and roam and explore the environment.</td>
<td>____ a</td>
<td>“You feel a little caged in, I must say.” [P7]</td>
</tr>
<tr>
<td>Sitting, passive enjoyment, and relax</td>
<td>Participants wanted to sit and relax in the environment, enjoying the view.</td>
<td>Add the possibility to sit on a recliner, bench, blanket on the ground, or in a boat on the water and have a picnic.</td>
<td>“I would like to be able to sit down and just sit and look and listen to birds singing, not do anything other than just relax.” [P5]</td>
</tr>
<tr>
<td>Authenticity and realism</td>
<td>In various ways, participants emphasized the importance of authenticity and realism in the VNE b.</td>
<td>There should be movement in the trees from the wind just as there is in the grass; otherwise, it is inconsistent.</td>
<td>“The trees look a little like trees on a model railway.” [P5]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“I think you need a rather high degree of realism.” [P7]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“It should be flowers that exist in reality.” [P1]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“I would have had difficulties [relaxing in an inconsistent environment]...because then I focus on that [the inconsistencies].” [P6]</td>
</tr>
</tbody>
</table>

aNot available.
bVNE: virtual natural environment.

https://humanfactors.jmir.org/2023/1/e40932
Table 4. Realization of participants’ preferences and their reactions.

<table>
<thead>
<tr>
<th>Implementations</th>
<th>Addressed preferences</th>
<th>Description</th>
<th>Reactions of the participants</th>
</tr>
</thead>
</table>
| Animals                          | Movement, life, and change; authenticity and realism                                     | Birds, fish, and a butterfly exhibiting natural-like behavior controlled by AI\(^a\) scripts | • “It’s considerably more natural, especially with the butterflies and birds.” [P6]  
• “It feels more alive, it doesn’t feel as artificial.” [P5]  
• “I thought it was positive with the butterfly and the birds, and the fish. It became more alive.” [P1]  
• “It must be a different country because such fish we don’t have here.” [P9]                                                                                   |
| Movement of the tree branches    | Movement, life, and change; authenticity and realism                                     | Animation of the tree branches to resemble movement from the wind             | • “It’s good that it moves a little; it feels considerably more natural.” [P5]  
• “The branches move a little. They didn’t do that last time. That’s nice. So that it’s something more that happens.” [P3]  
• “This branch over here moves but the trees over there do not...when you notice that, you feel that something is not right.” [P4] |
| Teleportation                    | Roaming and exploring                                                                   | Ability to instantly teleport oneself by aiming the handheld controller to an arbitrary point in the environment and pressing a button | • “This was a boost, absolutely. You become more active; you don’t simply stand and look around.” [P4]  
• “It feels more free.” [P8]  
• “It becomes considerably nicer than to be stuck in one place. You get more experiences.” [P5]  
• “It feels very artificial, that way to move. You are somewhat in a computer game context.” [P11]                                                               |
| Swivel-chair vehicle             | Roaming and exploring; sitting, passive enjoyment, and relaxation                       | Ability to drive in the environment while sitting in a swivel chair, press a controller button to instigate propulsion, and steer by turning oneself in the desired direction | • “This is an amazing feeling. It feels like the chair is moving. This was cool.” [P4]  
• “Here I could move where I wanted, and see that I moved.” [P8]  
• “It is no problem to handle this.” [P7]  
• “I think it’s very awesome, but it is not natural for me to move like this [because I’m not a wheelchair user].” [P5]  
• “It doesn’t go very fast. Can you increase the speed a little?” [P11]                                                                                         |
| Jetty                            | Roaming and exploring; movement, life, and change                                       | A jetty to accommodate exploring the lake and observing the fish             | • “It was very interesting to go out on the jetty and look down into the water.” [P9]  
• “It’s fun to walk out on the jetty; you see the fish better from the jetty.” [P7]  
• “I think this is rather fascinating, to stand and look down.” [P11]                                                                                         |

\(^a\)AI: artificial intelligence.
The Third Iteration

Table 5 presents the participants’ preferences and suggestions that were expressed during the think-aloud sessions of the previous iteration (second iteration). Preferences that emerged included bodies of water; the ability to look at, study, or inspect things; a rich diversity of activities; and human presence or activity.

We proceeded to analyze how the participants’ new preferences could be realized and made additions and changes to the prototype accordingly. Table 6 presents these implementations, the preferences they were meant to address, and some of the participants’ reactions that were gathered during the subsequent think-aloud sessions that concluded the third iteration. As shown in Figure 7, the implementations consisted of a small rowboat and a peninsula (Figure 7A), a sunken boat (Figure 7B), and an apple tree (Figures 7C and Figure 7D).

Figure 8 shows the IMI (Figure 8A) and SUS (Figure 8B) scores for the third iteration. The IMI scores indicated that the participants found the experience highly interesting and enjoyable and did not feel much pressure or tension and that their view of its value and usefulness to them was neutral to high. The median SUS score was just above excellent on the adjective scale by Bangor et al [36], which again indicated that, despite the increased complexity of use that was introduced with the rowboat and the apple tree, the participants viewed the prototype as highly usable.
<table>
<thead>
<tr>
<th>Preferences</th>
<th>Description</th>
<th>Participants’ suggestions</th>
<th>Illustrative quotes</th>
</tr>
</thead>
</table>
| Water                        | Participants expressed a preference for bodies of water and for being in or on the water. | Add the possibility to ride in or drive a boat such as a rowboat or canoe.                | • “I was actually delighted by the lake, especially with the translucency [so that] you could see the fish, and that you could drive out in it.” [P11]  
  • “I think you could preferably have a boat to step into...and then you could glide out on the lake.” [P1] |
| Interactivity                | Participants had many diverse preferences for activities in the VNE\(^a\). | Add the ability to pick things such as flowers, mushrooms, apples, or shells; go fishing, canoeing, and mountain climbing; and read a book, grill, swim, fly a kite, and play with their grandchildren. | • “You can imagine an apple tree, you pick apples.” [P4]  
  • “One should be able to pick some flowers.” [P3]  
  • “If you go out on the jetty, you could fish.” [P9] |
| Look at, study, or inspect things | Participants wanted to look at, study, or inspect interesting things in the environment. | Provide the possibility to look down into the water and see things such as clams, crabs, aquatic plants, or just the bottom; inspect birds closely; and look through binoculars. | • “[Y]ou stop if there is something that you find interesting...and you think ‘exciting,’ and I want to see what that looks like...you want to inspect it closer,” [P10]  
  • “That there was something on the bottom, to look [at] and contemplate, fish, and it can be whatever, rocks, clams.” [P8] |
| Human presence or activity   | Participants thought that one should be able to see or hear humans, human activity, or traces thereof. | People who are visibly present in the distance, walking by, swimming, or working in a garden; occasional sounds from agriculture, forestry, or a car in the distance; an airplane in the sky; and items or structures that reveal human presence, such as a bench, a fireplace, an old bicycle, or an old boat | • “There are no people around. It is very empty of people.” [P1]  
  • “[A] stone bench over by the beach somewhere, that you can imagine where people have sat and enjoyed themselves.” [P7]  
  • “You could see some pollution, an old bicycle. Something you don’t expect to see.” [P6] |

\(^a\)VNE: virtual natural environment.
Table 6. Realization of participants’ preferences and their reactions.

<table>
<thead>
<tr>
<th>Implementations</th>
<th>Addressed preferences</th>
<th>Description</th>
<th>Reactions of the participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small rowboat</td>
<td>Water; look at, study, or inspect things; interactivity; human presence or activity</td>
<td>A small rowboat that a user could enter and drive on the water</td>
<td>• “It was very nice, cozy in some way, and as if I had rowed out myself. It feels natural.” [P13]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• “It’s rather fascinating to look down into the water. It looks rather true to nature. There comes a little fish. It moves like a fish should.” [P4]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• “It’s a little weird that it moves without you rowing. It feels like you had had a small outboard motor.” [P11]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• “The reeds should move out of the way [when driving the boat over them].” [P6]</td>
</tr>
<tr>
<td>Sunken boat</td>
<td>Look at, study, or inspect things; human presence or activity</td>
<td>A sunken boat that could be discovered by driving past in the rowboat</td>
<td>• “It’s exciting because you didn’t see that it was a boat until you got closer.” [P9]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• “It’s good that things happen; that there are things...to discover...I think [it] makes it a little more interesting.” [P10]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• “You see that, gee, here is something. I must inspect it further.” [P5]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• “It also provides the feeling that there are people.” [P11]</td>
</tr>
<tr>
<td>Peninsula</td>
<td>Realism and authenticity; roaming and exploring</td>
<td>Reshaping of the landscape to form a peninsula in the lake; adding more trees and flowers with the intention of making the environment look more authentic and natural and more interesting to explore</td>
<td>• “[I]t looks much more natural, trees and such. It’s not as artificial as the first time.” [P5]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• “It became a more intimate landscape. It was like a desert before.” [P2]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• “I think it [the lake] is more natural now. It looked small and landscaped in the beginning.” [P1]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• “It is considerably more alive, more to discover.” [P4]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• “I don’t experience the ground as natural.” [P7]</td>
</tr>
<tr>
<td>Apple tree</td>
<td>Interactivity; human presence or activity</td>
<td>An apple tree from which users could pick apples and place them in a basket</td>
<td>• “This is a lot of fun.” [P12]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• “I think it’s good that you can do something; that you can drive in the boat, pick apples, walk around a little; that it doesn’t become just a passive experience.” [P5]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• “You feel involved, active; that you can do something yourself.” [P1]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• “I affect something [in the environment]. It enhances the experience.” [P4]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• “You can imagine that there was a farm here before and that an apple tree remains.” [P2]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• “It becomes unrealistic because if I pick berries, I want to be able to eat them.” [P10]</td>
</tr>
</tbody>
</table>
Figure 7. Implementations of the third iteration from the point of view of the participants while in the virtual natural environment: (A) driving the rowboat, (B) discovering the sunken boat, and (C) and (D) picking apples from the apple tree and placing them in the basket.

Figure 8. (A) Intrinsic Motivation Inventory and (B) System Usability Scale results for the third iteration.
Final Analysis

Upon completion of the third iteration, we proceeded to analyze the latest round of think-aloud sessions. Table 7 presents both preferences that relate to the latest implementations and those that surfaced throughout the study that did not pertain to any particular iteration or implementation. Participants reflected that discovering traces of human activity triggered their imagination. Generally, participants preferred the environment to be familiar and relatable. Other ideas that surfaced recurrently throughout the study were the ability to eat or drink, that the sound should be more realistic and varied, and being able to touch and smell the environment and feel the wind.

Table 7. Participants’ preferences and suggestions in the think-aloud session in the third iteration.

<table>
<thead>
<tr>
<th>Preferences</th>
<th>Description</th>
<th>Illustrative quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discovering traces of human activity triggers the imagination.</td>
<td>Participants reflected that traces of human presence or activity made the environment more interesting and set their imaginations in motion. They wondered what events had taken place, which people had been there before, and what had happened to them.</td>
<td>“You imagine what has happened to the people who lived here.” [P9]</td>
</tr>
<tr>
<td></td>
<td>Add more traces to discover a cairn, an old pot, ruins of an old house, traces of a garden, an old well, a root cellar, or an item that someone had lost or dropped. Make it so things change in between sessions—someone picked all the apples or hauled away the shipwreck.</td>
<td>“It sets the imagination in motion...You begin to wonder. It’s positive compared to nature that is completely free from human traces.” [P3]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“The little shipwreck can be gone the next time. Then you discover that someone has taken care of it, because it’s not so good that it lies there and rots.” [P1]</td>
</tr>
<tr>
<td>Familiarity and reminiscence</td>
<td>Participants preferred the environment to be familiar and to be able to identify with it and relate to memories, such as from their childhood.</td>
<td>“Something that you recognize; an environment like you grew up in.” [P8]</td>
</tr>
<tr>
<td></td>
<td>Make it possible to carry out activities one did as a child.</td>
<td>“An older person wants to recognize themselves [in the environment].” [P7]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“It’s quite nice if you can identify [species] so you can say, ‘Wow, it’s a great tit, or a blue tit.’” [P3]</td>
</tr>
<tr>
<td>Eating or drinking and socializing</td>
<td>Participants expressed a desire to eat or drink something, potentially in the form of a social situation such as having coffee together with others present in the VNE at the same time.</td>
<td>“You could have coffee out here...Suppose that there...is one more [person] who in some way interacts...perhaps sets the table, pours coffee, says something perhaps, ‘Come now it’s time for coffee.’” [P11]</td>
</tr>
<tr>
<td></td>
<td>The possibility to have a picnic on a blanket on the ground, a wooden bench table, or a cafeteria</td>
<td>“I can imagine sitting on one of these green slopes and having coffee or a little picnic.” [P14]</td>
</tr>
<tr>
<td>Sound</td>
<td>Participants thought that sounds that realistically should be there were missing and that there should be more variation.</td>
<td>“I would have liked to hear splashing [while driving the boat]. It would have strengthened the illusion.” [P5]</td>
</tr>
<tr>
<td></td>
<td>Sounds from the wind, rustling when walking over fallen leaves, and splashing of water when driving the boat; sounds that bring variation: a cuckoo or pigeon, a car in the distance, or people walking</td>
<td>“You could hear a sound in the background; a tractor, or a boat on the lake, some momentary sounds.” [P2]</td>
</tr>
<tr>
<td>Other modalities</td>
<td>Participants wanted to experience the environment through additional senses.</td>
<td>“You would like to touch them [the trees] even though you know that there is nothing. If I get really close to it, like this, you want to touch it. But then there is nothing.” [P10]</td>
</tr>
<tr>
<td></td>
<td>To feel the breeze and the warmth of the sun; touching things, trees, rocks, or apples; smelling things, flowers, or coffee; having feet lowered into a bucket of water while being in the virtual water</td>
<td>“Wind is probably the most important that you can feel...so that you feel like you are outside.” [P1]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“If I’m out in nature there are often smells. That I don’t sense here.” [P11]</td>
</tr>
</tbody>
</table>

aVNE: virtual natural environment.

Observations of Usability

Although the SUS scores indicated high usability throughout the study, in our observations of the participants exploring the environment, we noticed that some had difficulties finding the correct buttons on the controller for teleporting and driving. Sometimes, they teleported by mistake by accidentally pushing the teleport button. The task of approaching the apple tree and ending up in a location convenient for picking apples seemed difficult at times. This was because the participants often came too close to the tree or even moved inside it. The participants often flinched when they came too close to the branches.

VR Symptoms

The Virtual Reality Symptom Questionnaire revealed very few symptoms, with means of <1 on a scale from 0 to 6 on every symptom measured in all 3 iterations. However, there were isolated medium-high scores reported: a score of 3 for dizziness.
Principles of a Meaningful VNE

Overview
As a final step in the process, we took a bird’s-eye view of all the participants’ preferences to reveal recurring threads of meaning and condense the findings into applicable principles. By reflecting on the patterns and interrelationships of the preferences, 3 main principles emerged—realness, interactivity, and relatedness. These could be considered by designers of future VNEs. However, they need to be tested and potentially revised in future studies.

Realness
Realness refers to how complete the experience is in terms of presence, realism, and believability. A VNE should provide a sense of presence (ie, a sense of being in place). This can be accomplished by providing (approximate) real-world responses to actions [40]. Sideways head movements, for example, should allow the user to look behind objects, and touching objects should optimally provide haptic feedback:

It would be very interesting to sense that feeling [touch the rock]...For me, it would be very positive. Because then I am absolutely in nature. [P6]

A VNE should provide believability. By this we mean a sense of a complete world that does not end behind a backdrop—a world in which there are interesting things to discover, such as animal and plant life, traces of human activity, details that reveal natural processes, and distant sounds. We propose that these characteristics can contribute to a feeling of a complex living world that continues beyond the reach of one’s eye and stretches back into a historical past:

It’s a very static environment. I must say that I don’t feel like I’m out in nature for real, more like I’m standing on a stage with backdrops. [P11]

It’s more alive, just that little butterfly alone contributes a lot, I must say. [P1]

[Here has been a house, and there you can see an old apple tree... it triggers the imagination, that someone has lived here and how were they doing? How were they able to live here in the middle of nowhere? Such things can also contribute a little. [P5]

It provides realism to it all. I mean, it’s certainly not unusual that there lies an old miserable-looking boat by the shoreline. [P14]

The trees look too nice... there are none [branches] that has been broken off by the wind and such. [P5]

A VNE should provide a sense of trueness to reality, such as accuracy in rendition and the combination and behavior of natural and human-made elements:

It should be flowers that exist in reality. [P1]

I don’t expect to see an apple tree in a pine forest. [P11]

You stop enjoying it, lose interest in it... when you go into the details and it is incorrect. Then it’s maybe a little lesser of an experience. [P9]

Interactivity
In addition to basic VR sensory-motor contingencies—including moving one’s body, head, and eyes to change gaze direction and looking behind or under things—a VNE should preferably provide the possibility to engage in activities (eg, exploring the environment and picking mushrooms). Activities should be congruent with the users’ capabilities, interests, and identities. Care should be taken to also accommodate for passive activities (eg, to just sit and relax looking at things). As previously stated, user actions should elicit responses that are as close to reality as possible:

To be able to do different things, walk around, look down into the water properly, drive the boat, pick apples. They are positive elements in it compared to the first time, [which was] just a quiet and rather flat environment. Suddenly a lot of things happen, you can do a lot of things. [P5]

It was a lot of fun to pick the apples and put them in the basket. It was an extra-experience... because you do something, and you see that it works. But, if you drop an apple, normally, “thud” it says, but these were very soft. [P13]

Relatedness
A VNE should enable a user to identify with it and relate to memories (eg, the environment should be familiar, and the vegetation and animals should be of species that one can recognize):

[It should be an environment] that you recognize yourself in, that you can relate to. [P3]

Where I used to walk when I was a child, for example, I think would be a very nice experience... Because then you get it related directly to yourself. [P13]

I have some good friends who have a family farm by a small lake... I can imagine that if they saw this, it would evoke memories in them in their old age; but to evoke such memories in me, it should be by the sea. [P10]

To me, when it comes to nature... you relate to the memories that you have from nature, and if it matches better, it feels like it’s more natural. [P8]

Discussion
Principal Findings
This study was conducted to investigate older adults’ preferences and ideas regarding VNEs and how these can be realized. To our knowledge, this is the only study in which such preferences for VNEs were collected through an iterative participatory design process using an inductive approach. Our results suggest the preferences outlined in Textbox 1.

It should be noted that preferences for activities were very diverse among our participants. Overall, we propose that one
should pay attention to multiple choices of activities and variations of features. Ideally, a VNE should be tailored to the individual. The participants’ responses to the questionnaires and reactions to our implementations suggest feasible ways to realize some of the preferences for VNEs that emerged during this study (Textbox 2).

**Textbox 1. Participant preferences.**

<table>
<thead>
<tr>
<th>Preferences for realness</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Trueness to reality in the rendition and behavior of natural and human elements (e.g., accuracy in prevalence and combination of species)</td>
</tr>
<tr>
<td>• Real-world responses to actions, such as accurate audial and visual feedback (e.g., displacement of objects and sound of impact)</td>
</tr>
<tr>
<td>• Extended range of sensory modalities such as touch, smell, and temperature</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Preferences for interactivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>• The possibility to roam and explore the environment and inspect interesting things such as natural and human-made elements</td>
</tr>
<tr>
<td>• The possibility to engage in activities such as picking mushrooms, taking a boat ride, or having a picnic</td>
</tr>
<tr>
<td>• The possibility to sit and just relax observing the environment</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Preferences for relatedness</th>
</tr>
</thead>
<tbody>
<tr>
<td>• A familiar, relatable environment in which one can recognize oneself</td>
</tr>
<tr>
<td>• Being able to recognize and identify species and natural features</td>
</tr>
<tr>
<td>• Being able to relate to memories</td>
</tr>
</tbody>
</table>

**Textbox 2. Feasible ways to realize participant preferences.**

<table>
<thead>
<tr>
<th>Realizations of realness</th>
</tr>
</thead>
<tbody>
<tr>
<td>• The occurrence of animals (e.g., birds, fish, and insects) and the movement of vegetation and other natural features (e.g., from wind) make the environment appear more alive and real.</td>
</tr>
<tr>
<td>• Traces of human activity trigger the imagination, allowing users to picture people and past events in the environment. This contributes to their perception of the environment as a real place.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Realizations of interactivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>• The possibility to roam and explore the environment enhances the experience of a virtual natural environment. It can activate and provide a sense of increased freedom to users.</td>
</tr>
<tr>
<td>• Teleportation and forward propulsion with seated swivel chair–directed steering are feasible methods for roaming and exploring the environment. However, they can be perceived as unnatural in their own ways.</td>
</tr>
<tr>
<td>• Activities that involve some manipulation of the environment, such as picking apples, can enhance the experience, making users feel involved and active. However, unrealistic mechanics of activities may be unfavorable for the experience.</td>
</tr>
</tbody>
</table>

Some of our results showed that participants wanted to be active in the VNE. They wanted to explore, inspect, and interact with the environment (e.g., pick the flowers, touch the rocks, and feel the water). At the same time, they wanted their actions to be true to reality. This is not so surprising as perception requires action [41]. For example, shifting one’s head sideways enables one to see behind things. According to Slater et al [40], VR works by providing real-world responses to actions. That is how one acquires a sense of presence, of being in place in the virtual environment. If one performs an action and there is no response or the response is too far from reality, one will momentarily break presence. Thus, consistent with our findings, we propose that the possibility of interaction (that approximates real-world interaction) is important in VNEs. Although this proposal may seem obvious, many studies of VNEs use technologies that feature minimal or no interaction, such as 360° video [26], which does not support sideways head movements. However, 360° video has other advantages, such as the ability to capture existing environments, which can make the participants’ preference for relatedness (i.e., familiar environments and the possibility to relate to memories) more economical and feasible to realize. Orr et al [20] used 360° video to provide VR experiences of local beaches to older adults with mild to moderate cognitive or memory impairment. They found that familiarity with the environments brought the participants enjoyment in identifying places and relating to memories. They also found indications that the VR provided was “sufficiently credible” and that participants were able to acquire a sense of presence. However, the presence of other (filmed) people in the 360° videos provoked in the participants a desire to interact, which, naturally, was impossible. A recent study of VNE use by patients with breast cancer advised that future studies should “focus on activities that encourage connection with nature (rather than simply exposure to nature)” [42].
The biophilia hypothesis argues that humans have a biologically based need for a sense of belonging to the natural world. This connectedness with nature is instrumental for human well-being according to the hypothesis [25]. One can speculate that our participants’ preference for being able to identify with and relate to the environment is partly an expression of nature connectedness, and interaction with the environment has the potential to promote connectedness and relatedness. Other studies have measured nature connectedness in VNEs. One study [16] found higher nature connectedness in participants using an immersive VNE than in those watching a corresponding natural environment on a television screen. Another study [21] suggested that nature connectedness mediated positive affect in older adults using an immersive VNE.

On the basis of the participants’ preferences and reactions to our implementations, we propose that a VNE for older adults should be true to reality in rendition and behavior and that traces of human activity and natural processes can promote the perception of it as a real place. This is very reminiscent of what Slater et al [40] present as a contributing factor to plausibility (Psi), which is the illusion that the events happening in the virtual environment are actually taking place. In addition to trueness to reality in appearance and behavior, Psi relies on the virtual environment in some way acknowledging the user (eg, that a virtual human character plausibly responds to an action by the user). Although this study did not involve social interaction with virtual human characters, we found that our proposal is congruent with the presentation of Psi by Slater et al [40].

Limitations and Future Research

The focus of this study leaned heavily toward gathering the participants’ ideas, preferences, and suggestions and revealing unforeseen problems and considerations in the process of designing a VNE. This focus was a determining factor in the choice to include the think-aloud protocols in the study design. There is a risk that the extensive recurring think-aloud sessions generated bias within the participants, which could have affected the scoring of the questionnaires. It is not difficult to imagine that the participants may have adopted a positive stance as they had someone who was listening carefully to them and was genuinely interested in their opinions, as well as experiencing firsthand that their opinions had an impact on the design of an important innovation. As the participants generally had very little experience with VR (Table 2), its novelty could have contributed to the high IMI scores. The findings of this study need to be tested and potentially revised in future studies.

As is evident from the inclusion and exclusion criteria and the results of the background questionnaires, the participants in this study were relatively healthy and mobile and had access to natural environments. Thus, they differed in many ways from the proposed group that this study is aimed at, namely, residents of residential care facilities who have limited or no access to natural environments because of ill health. The lower scores the participants gave on the value and usefulness dimension of the IMI questionnaire compared with the other dimensions may be a testament to this fact—they simply may not have found the VNE very useful as they were able to visit real natural environments. This study was also carried out in a controlled laboratory environment and so did not consider the complexity of residential care facilities and the network of different stakeholders involved. Consequently, the next step would be to continue these research efforts with residents and staff at care facilities in the real world.

The HTC Vive has a setting to adjust the distance between the lenses to match a user’s interpupillary distance (IPD). In some cases, an inaccurate IPD setting may cause visual distortion and discomfort [43]. Ideally, our test protocol should have included measuring each participant’s IPD and configuring their headset accordingly. However, to simplify the test procedure, we opted to set the IPD to 63 mm, which is the average in adults [44], and only adjusted the IPD setting for users if they reported discomfort or an unacceptable image quality. The participants’ reports of blurred vision may partly be explained by an incorrect IPD setting. Future study protocols should include measuring the participants’ IPD to improve the perceived image quality and reduce the risk of eye strain and discomfort.

Conclusions

This paper described an iterative user-centered design process for a VNE for older adults. We presented the participants’ preferences and ideas, how these were realized in the ongoing development of a VNE prototype, and how new implementations were received by the participants. We proposed 3 principles for VNEs for older adults: realness, interactivity, and relatedness. We also provided suggestions that can be considered by designers and researchers of VNEs for older adults. These include trueness to reality in terms of rendition and behavior; traces of human activity and natural processes that trigger the imagination and provide believability; the ability to roam, explore, and interact with the environment; and a familiar, relatable environment that evokes memories. VNEs should provide a diversity of content and activities to accommodate the heterogeneity in older adults’ preferences. We argued that non–VR-related theories of restorative natural environments may not be applicable in a VR context. This study can contribute to the development of a framework for designing VNEs for older adults.

Acknowledgments

This study was funded by the European Union Horizon 2020 research and innovation funding program under grant agreement 666773. The authors would like to thank all the participants in this study for giving their time and effort to test and share their ideas, preferences, suggestions, and opinions.
Authors' Contributions

Study conception, study design, material preparation, and analysis were conducted by the first author (RL) and the last author (MW). Software development and data collection were conducted by RL. The first draft of the manuscript was written by RL. All authors were involved in the discussions of the analysis, commented on previous versions of the manuscript, and read and approved the final manuscript.

Conflicts of Interest

RL and MW are planning to register a business to provide virtual reality services for older adult care.

References


34. Moerman G. Probing Behaviour in Open Interviews : A Field Experiment on the Effects of Probing Tactics on Quality and Content of the Received Information. Amsterdam: VU University; 2010.


Abbreviations

HMD: head-mounted display
IMI: Intrinsic Motivation Inventory
IPD: interpupillary distance
Psi: plausibility
SUS: System Usability Scale
VNE: virtual natural environment
VR: virtual reality
Empowering Researchers to Query Medical Data and Biospecimens by Ensuring Appropriate Usability of a Feasibility Tool: Evaluation Study

Christina Schüttler1*, PhD; Maria Zerlik2*, MSc; Julian Gruender3, PhD; Thomas Köhler4,5; Lorenz Rosenau6, MSc; Hans-Ulrich Prokosch1, PhD; Brita Sedlmayr2, PhD

1Central Biobank Erlangen, University Hospital Erlangen, Erlangen, Germany
2Institute for Medical Informatics and Biometry, Carl Gustav Carus Faculty of Medicine, Technische Universität Dresden, Dresden, Germany
3Chair of Medical Informatics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany
4Federated Information Systems, German Cancer Research Center, Heidelberg, Germany
5Complex Data Processing in Medical Informatics, Medical Faculty Mannheim, Mannheim, Germany
6IT Center for Clinical Research, University of Lübeck, Lübeck, Germany

*these authors contributed equally

Abstract

Background: The Aligning Biobanking and Data Integration Centers Efficiently project aims to harmonize technologies and governance structures of German university hospitals and their biobanks to facilitate searching for patient data and biospecimens. The central element will be a feasibility tool for researchers to query the availability of samples and data to determine the feasibility of their study project.

Objective: The objectives of the study were as follows: an evaluation of the overall user interface usability of the feasibility tool, the identification of critical usability issues, comprehensibility of the underlying ontology operability, and analysis of user feedback on additional functionalities. From these, recommendations for quality-of-use optimization, focusing on more intuitive usability, were derived.

Methods: To achieve the study goal, an exploratory usability test consisting of 2 main parts was conducted. In the first part, the thinking aloud method (test participants express their thoughts aloud throughout their use of the tool) was complemented by a quantitative questionnaire. In the second part, the interview method was combined with supplementary mock-ups to collect users’ opinions on possible additional features.

Results: The study cohort rated global usability of the feasibility tool based on the System Usability Scale with a good score of 81.25. The tasks assigned posed certain challenges. No participant was able to solve all tasks correctly. A detailed analysis showed that this was mostly because of minor issues. This impression was confirmed by the recorded statements, which described the tool as intuitive and user friendly. The feedback also provided useful insights regarding which critical usability problems occur and need to be addressed promptly.

Conclusions: The findings indicate that the prototype of the Aligning Biobanking and Data Integration Centers Efficiently feasibility tool is headed in the right direction. Nevertheless, we see potential for optimization primarily in the display of the search functions, the unambiguous distinguishability of criteria, and the visibility of their associated classification system. Overall, it can be stated that the combination of different tools used to evaluate the feasibility tool provided a comprehensive picture of its usability.
usability evaluation; ontology; feasibility queries; user-centered design; clinical research informatics; user interface

Introduction

Background

The past decade has seen various projects aimed at making medical data and biological samples available for research. On a national level, the German Biobank Node (GBN) [1] pioneered biobanking, whereas the Medical Informatics Initiative (MII) [2] was able to establish infrastructure for processing and analyzing patient data from routine care by setting up data integration centers (DICs) at German university hospitals. In 2021, it was decided to merge these projects, which had previously run in parallel. The resulting project—Aligning Biobanking and Data Integration Centers Efficiently (ABIDE)—aims to harmonize technologies, regulations, committees, and governance structures of the 24 participating German university hospitals and their 25 biobanks to create a single point of contact for researchers searching for patient data and (associated) biospecimens. The central element will be a feasibility tool that researchers can use to query the availability of data and samples from routine care at the connected sites to determine the feasibility of their study project. The development of the tool should take into account that potential end users (laypersons) usually do not have specific knowledge regarding the execution of queries and that a too complex user interface, as found in, for example, expert tools such as ATLAS [3], should be avoided.

The ABIDE project benefits from previous work using the infrastructure of the DICs established within the MII. In addition, the ABIDE project takes advantage of the experience gained from the German Biobank Alliance (GBA) [4], which is coordinated by the GBN. Beyond this, the development work of the Network University Medicine COVID-19 Data Exchange Platform (CODEX) [5] project can be seamlessly integrated. In the CODEX project, based on prespecified requirements, a first test version of the envisaged feasibility tool (hereinafter referred to as feasibility tool v1) for simple queries has already been implemented and evaluated by potential end users regarding user-friendliness [6-8].

The feasibility tool v1 at the time allowed a simple querying of data elements based on the COVID-19–specific German Corona Consensus Data Set (GECCO) [9] and executing of federated queries on the Fast Healthcare Interoperability Resources (FHIR) servers at the MII DICs at the distributed sites. Data elements could be selected either via a free-text search field or a category tree and added as inclusion or exclusion criteria to a query. In addition, the criteria could be linked using Boolean operators. The usability analysis of the feasibility tool v1 showed that the previous developments were perceived as positive by users [8]. In particular, users found the intuitive operating concept convincing.

Nonetheless, some usability problems were uncovered. Among the points noted were a need for clearer visualization of the subdivision of inclusion and exclusion criteria, a uniform display of linking using Boolean operators, and the ability to search for synonyms. In addition, a function was desired to save a created query and continue editing it later or to archive sent queries together with the results. These and other functionalities were the focus of the development of an improved version of the feasibility tool in the ABIDE project (hereinafter referred to as feasibility tool v2) as additional requirements. In addition, focus was placed on the integration of the temporal restriction of criteria, grouping of criteria, and representation of their temporal relationship to each other, which was defined as an additional technical development goal for the ABIDE project. Another priority was the extension of the searchable data set. This was intended to expand the underlying ontology to the entire core data set of the MII [10], including biospecimens, so that it would no longer be limited to the GECCO.

In this way, the entire patient collective of the participating university hospitals can be considered in future study cohorts by means of appropriate feasibility queries. Furthermore, the integration of the feasibility tool v2 into the German Research Data Portal for Health (Forschungsdatenportal für Gesundheit [FDPG]) [11] will allow researchers to coordinate their research centrally via 1 platform.

The planned implementation was tested during development using a simulation prototype and supplementary mock-ups. On the basis of feedback, a revised version of the feasibility tool (v3) will be created, which can then serve the development team as a reference for the final programming.

Objectives

The objectives of the study were as follows: (1) an evaluation of overall user interface usability, (2) the identification of critical usability issues, (3) comprehensibility of the underlying ontology operability, and (4) analysis of user feedback on additional functionalities. From these, recommendations for quality-of-use optimization, focusing on more intuitive usability, were derived.

Methods

Study Design

We conducted an exploratory usability test consisting of two main parts:

1. The thinking aloud method, in which test participants express their thoughts aloud throughout their use of the tool, was complemented by a quantitative questionnaire.
2. The interview method was combined with supplementary mock-ups to collect users’ opinions on possible additional features.

The participants tested the feasibility tool v2 on the web from their workplace. Neither randomization into intervention and control groups nor blinding took place.
Ethics Approval

The ethics committee at the Friedrich-Alexander-Universität Erlangen-Nürnberg approved the study (21-420-S).

Recruitment

The focus of the study was on the primary user group of the feasibility tool v2. These are researchers who have a research question and require a cohort with specific patient data or available biospecimens to address it. Professionals with a biobanking background and IT specialists with a research background were also recruited. This is because they are considered a secondary user group as it can be assumed that they will also use the tool (eg, to process internal queries). Recruitment was initiated and coordinated by the ABIDE project management, and potential study participants were contacted through project staff at each site. One prerequisite was that the participants should have had no prior experience with the tool to be tested. This prevented an overlap with those who tested the first prototype. In accordance with the requirements of the study protocol, a sufficient number of individuals were approached to achieve the sample size of at least 14 volunteers.

Description of the Feasibility Tool v2

The feasibility tool v2 was evaluated in January 2022. Compared with the first release, this version includes the core MII data set in addition to the GECCO. This enhancement means that the FDPG can ultimately serve as a central point of contact for people who want to check the Germany-wide availability of data and biospecimens from affiliated university hospitals to answer their research questions. In alignment with study protocols, in which exclusion and inclusion criteria are usually formulated for research questions, the interface of the feasibility tool v2 was designed to be structured accordingly (Figure 1).

Criteria that are relevant for the study or should be avoided can be searched for in the respective areas using either a free-text search or a category tree (Figure 2) and selected. After the initial selection of the criterion, a pop-up window opens offering the possibility to add further restrictions (Figure 3).

In addition to criterion-specific restrictions (eg, specification of a value range or the localization of the biospecimen), a temporal constraint is possible. The possibility to link the selected criteria using Boolean operators is offered as soon as the criteria have been finally added to the query. Once the desired query has been formulated, it can be executed (Figure 4).

As soon as the search query is processed, the result is displayed in the upper area under Number of patients. The Details option provides an overview of the breakdown of the cumulative result, although the data-providing hospitals remain anonymous.

![Figure 1. Search interface of the feasibility tool v2 of the Aligning Biobanking and Data Integration Centers Efficiently project.](image-url)
Figure 2. Search options via free-text search or category tree.

Figure 3. Pop-up window with the possibility to specify selected criteria.
Study Flow
Interested participants were enrolled in the study after being recruited and provided with detailed information, including an informed consent form and a privacy statement. Upon receipt of the signed forms, an appointment was made to conduct the evaluation, which lasted approximately 60 minutes. After attending a brief welcome session and having been provided an overview of the study, the participants had to solve 3 tasks as part of an exploratory usability walk-through. The test leader protocollled the testing and the comments of the participants in a structured form. After the participants had completed the test tasks, we collected information regarding usability, demographic aspects, expertise, and so on, using the web-based survey tool SoSci Survey [12]. Subsequently, participants were able to provide their input on the various additional functions presented using mock-ups. Feedback on the acceptance and added value of these possible implementations was collected using a structured interview.

For backup reasons, the entire session was captured on Zoom (Zoom Video Communications, Inc) [13] using the videoconferencing platform’s recording function and stored in a password-protected cloud folder.

Instruments

Tasks
The evaluation team had compiled 2 test tasks themselves to be able to cover the entire range of functions of the feasibility tool v2 as far as possible. Care was taken to ensure that these tasks reflected realistic requests and varied in their degree of complexity. Moreover, a third task was formulated based on a real-world request submitted during an MII workshop. While carrying out the tasks, the test participants were encouraged to express their thoughts aloud according to the thinking aloud method [14,15]. The aim of this method was to gather immediate feedback on the strengths and weaknesses of the tool. In addition, suggestions for improvement, if any, were noted. The correctness of the tasks was evaluated by checking whether all criteria were correctly selected and linked and led to the required query. A scoring system was used to determine the number of points for each task performed. The test tasks can be found in Multimedia Appendix 1.

Questionnaires
After completing the tasks, the test participants were asked to describe their immediate impression of the feasibility tool v2 and, in particular, to list positive and negative design aspects as well as make suggestions for improvement. Subsequently, they were asked to assess the usability of the query tool using the System Usability Scale (SUS). According to Brooke [16], the SUS is a standardized and validated questionnaire that allows a quantitative assessment of the usability of the tested systems. In addition to the SUS questions on general usability, 4 more questions focused on the usability of the category search. Furthermore, the test participants were asked to answer questions regarding personal details, expertise, and experience. The questionnaires that were used in the evaluation can be found in Multimedia Appendix 2.

Interview and Mock-ups
A final interview block [17] served to determine user preferences regarding the implementation of new functions and whether this was congruent with the intended implementation. Mock-ups were created for the additional functions groups and temporal dependencies based on exemplary queries (Figures 5 and 6).
The corresponding interaction path was demonstrated to the test participants by the test leader for illustration purposes. On the basis of these mock-ups, the participants were asked to assess whether they perceived the approach as intuitive and, if not, what navigation path they would have expected. For the representation of temporal dependencies among the criteria or criteria groups, in the sense that, for example, conventional therapy was provided before an interventional procedure, the
participants were asked whether they would see added value in this and how functional such a representation would have to be (in terms of the number of criteria that would have to be linked).

Finally, the necessity to represent nested criteria—in terms of linking a criterion with another criterion, such as the International Classification of Diseases, Tenth Revision (ICD-10), diagnosis D43 (neoplasms of uncertain or unknown behavior of the brain and central nervous system) combined with the International Classification of Diseases for Oncology, version 3 (ICD-O-3), morphology 9383/1 (subependymoma)—was discussed with the participants. To find out the preferences of users, the test participants received an exemplary query to illustrate the problem. Although in its current state of development the tool did not offer the possibility to formulate this query in a single query, the test participants were motivated to express which approach they would have intuitively chosen or which functionality they would have expected to be able to formulate the query correctly.

Data Analysis

Analysis of the Thinking Aloud Protocols

After the test sessions, the task processing protocols were checked for completeness, supplemented if necessary, and electronically documented. All positive and negative aspects of the tool were extracted from the protocols. Three usability and ontology experts categorized the problems separately as usability related or ontology related. The consensus decision was documented in a list. In cases where a sharp distinction between usability-related problem and ontology-related problem was not possible, these were grouped in a separate cluster. The negative aspects were additionally rated by 2 experts using the severity scale developed by Nielsen et al [18].

Task Success

The correctness of task processing was both evaluated globally and differentiated for the respective task steps using a self-developed scoring system (Multimedia Appendix 3). We analyzed the mean score achieved across all participants, the SD, and the accuracy rate in percentages.

Analysis of the Web-Based Questionnaire

Regarding the SUS, we applied a quantitative evaluation using the scoring method formulated by Brooke [16]. The responses provided to the additional questions related to the criteria search were summed up per item. For questions regarding the participant, a descriptive evaluation (frequencies, mean values, and SDs) was performed.

Analysis of the Interview on Additional Features

Analogous to the thinking aloud protocols, the feedback obtained during the interviews regarding the additional functionalities was recorded and documented electronically. The statements were subjected to a descriptive qualitative content analysis.

Results

Sample Description

The study cohort consisted of 22 test participants from 14 ABIDE partners. This corresponds to 92% (22/24) of the potential participants approached and thus comfortably exceeds the planned sample size of 14 test participants. The majority of the study cohort was composed of the younger age groups 25 to 34 years (8/22, 36%) and 35 to 44 years (11/22, 50%). Of the 22 participants, 9 (41%) were male, and 13 (59%) were female; in terms of profession, 7 (32%) were researchers, 4 (18%) had a biobanking background, 8 (36%) were IT professionals with a research background, and 3 (14%) were from other groups or did not specify. Work experience averaged 4.65 (SD 5.34) years. Participants declared no (10/22, 45%) or only some (12/22, 55%) prior experience with feasibility queries; prior experience with similar systems was reported by only 9 (41%) of the 22 participants. Whereas the test participants rated their IT knowledge as at least medium (medium: 9/22, 41%, and high: 15/22, 68%), the ratings on medical knowledge ranged from very low (2/22, 9%) to rather low (5/22, 23%) and medium (8/22, 36%) to rather high (7/22, 32%). The detailed characteristics of the study cohort are shown in Table 1.
### Table 1. Detailed characteristics of the study cohort.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age group (years), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>25 to 34</td>
<td>8 (36)</td>
</tr>
<tr>
<td>35 to 44</td>
<td>11 (50)</td>
</tr>
<tr>
<td>45 to 54</td>
<td>2 (9)</td>
</tr>
<tr>
<td>55 to 64</td>
<td>1 (5)</td>
</tr>
<tr>
<td><strong>Sex (observed, not asked), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>9 (41)</td>
</tr>
<tr>
<td>Female</td>
<td>13 (59)</td>
</tr>
<tr>
<td><strong>Profession, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Researcher</td>
<td>7 (32)</td>
</tr>
<tr>
<td>Professional with biobanking background</td>
<td>4 (18)</td>
</tr>
<tr>
<td>IT professional with research background</td>
<td>8 (36)</td>
</tr>
<tr>
<td>Other</td>
<td>2 (9)</td>
</tr>
<tr>
<td>Not specified</td>
<td>1 (5)</td>
</tr>
<tr>
<td><strong>Work experience (years), mean (SD)</strong></td>
<td>4.65 (5.34)</td>
</tr>
<tr>
<td><strong>Prior experience with feasibility queries, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>10 (45)</td>
</tr>
<tr>
<td>Some</td>
<td>12 (55)</td>
</tr>
<tr>
<td><strong>Prior experience with similar systems, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>13 (59)</td>
</tr>
<tr>
<td>Yes</td>
<td>9 (41)</td>
</tr>
<tr>
<td><strong>IT knowledge, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>9 (41)</td>
</tr>
<tr>
<td>High</td>
<td>13 (59)</td>
</tr>
<tr>
<td><strong>Medical knowledge, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Very low</td>
<td>2 (9)</td>
</tr>
<tr>
<td>Rather low</td>
<td>5 (23)</td>
</tr>
<tr>
<td>Medium</td>
<td>8 (36)</td>
</tr>
<tr>
<td>Rather high</td>
<td>7 (32)</td>
</tr>
</tbody>
</table>

### Task Success

The effectiveness analysis (completeness and accuracy) showed that no participant managed to solve all the tasks correctly (in the sense of matching the model solution). Task 1a was successfully completed by half of the test participants (11/22, 50%). Task 1b displayed the best performance with a success rate of 100%. In task 2, of the 22 participants, 14 (64%) obtained the correct result. By contrast, only 1 (5%) of the 22 participants was able to solve task 3.

The accuracy analysis of the partial steps that had to be processed within the assignments based on the scoring system is presented in Table 2.

### Table 2. Task success according to the scoring system.

<table>
<thead>
<tr>
<th>Task</th>
<th>Maximum possible score</th>
<th>Mean score achieved (SD)</th>
<th>Accuracy rate, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1a</td>
<td>8</td>
<td>7.23 (1.28)</td>
<td>90.37</td>
</tr>
<tr>
<td>Task 1b</td>
<td>1</td>
<td>1.00 (0.00)</td>
<td>100</td>
</tr>
<tr>
<td>Task 2</td>
<td>8</td>
<td>7.64 (0.48)</td>
<td>95.50</td>
</tr>
<tr>
<td>Task 3</td>
<td>5</td>
<td>3.32 (0.87)</td>
<td>66.40</td>
</tr>
</tbody>
</table>
Of the maximum possible 8 points in task 1a and task 2, participants obtained an average of 7.23 (SD 1.28) points and 7.64 (SD 0.48) points, respectively. This corresponds to a success rate of 90.37% and 95.50%, respectively. Task 3 could only be completed correctly with an accuracy of 66.40%. With a maximum of 5 possible points, this corresponds to an average of 3.32 (SD 0.87) points scored. In task 1a, the major source of error was the choice of diagnosis (8/22, 36%). Instead of choosing “Essential (primary) hypertension,” participants often selected another characteristic containing the term “hypertension” (e.g., “Hypertension [hypertensive disease]”).

The same potential for error was present in task 3 for both criteria (“Vancomycin” [selected by 18, 82% of the 22 participants] and “treated in intensive care” [selected by 7, 32% of the 22 participants]) being searched. Less frequently, errors occurred because of an incorrect AND or OR used to link the criteria (5/22, 23%) or when entering time constraints (5/22, 23%).

**Global Assessment of Usability (SUS Score)**

Textbox 1 shows the respective mean scores of the SUS items. The SUS score of the feasibility tool v2 calculated across all participants was 81.25 (SD 13.42) on a scale of 0 to 100.

<table>
<thead>
<tr>
<th>System Usability Scale item and mean (SD) values</th>
</tr>
</thead>
<tbody>
<tr>
<td>I think that I would like to use this query tool frequently: 4.6 (0.5)</td>
</tr>
<tr>
<td>I found the query tool unnecessarily complex: 1.6 (1.0)</td>
</tr>
<tr>
<td>I thought that the query tool was easy to use: 4.3 (0.8)</td>
</tr>
<tr>
<td>I think that I would need the support of a technical person to be able to use this query tool: 1.9 (1.0)</td>
</tr>
<tr>
<td>I found the various functions in this query tool were well integrated: 4.1 (0.7)</td>
</tr>
<tr>
<td>I thought there was too much inconsistency in this query tool: 1.8 (0.9)</td>
</tr>
<tr>
<td>I would imagine that most people would learn to use this query tool very quickly: 4.3 (0.8)</td>
</tr>
<tr>
<td>I found the query tool very cumbersome to use: 1.5 (1.0)</td>
</tr>
<tr>
<td>I felt very confident using the query tool: 3.8 (0.9)</td>
</tr>
<tr>
<td>I needed to learn a lot of things before I could get going with this query tool: 1.8 (0.7)</td>
</tr>
</tbody>
</table>

The mean SUS score of test participants classified as IT professionals with research background was 88.75 (SD 8.00), which, in comparison with the mean SUS scores of the primary user groups researchers (mean 78.13, SD 9.00) and professionals with biobankingbackground (mean 70.00, SD 13.46), was slightly higher.

The evaluation of the findability of criteria—based on the questions formulated in addition to the SUS scores—by the study participants indicates that the search for criteria was perceived as easy. Participants found that searching via the category tree tended to be more difficult than via the free-text search. More than half of the participants (14/22, 64%) had the impression that they could easily find the relevant criteria to solve the test tasks. Figure 7 shows the rating of the 4 additional items.

**Usability Aspects Identified**

**General Aspects**

The analysis of the thinking aloud protocol revealed that the majority of the participants (13/22, 59%) assessed the user interface of the feasibility tool v2 as simple to use and intuitive.

Searching for criteria using the free-text search was frequently emphasized as a helpful feature. Moreover, the clarity of the user interface and visual separation of the inclusion and exclusion criteria were highlighted as particularly positive. The switch button that makes it easy to change AND to OR was considered a well-integrated solution.
In addition to the positive aspects, 39 usability problems were identified and classified using the severity scale developed by Nielsen et al [18] as follows: 5 (13%) were classified as usability catastrophes, 8 (21%) as major usability problems, 12 (31%) as minor usability problems, and 14 (36%) as cosmetic problems.

Among the 5 usability catastrophes was that the free-text search bar was not easily located since the free-text input fields are grayed out indicating inactivity. In addition, the identification of relevant criteria in the results list of the free-text search was partly perceived as difficult, first because of the missing labeling of the code type and second because of the absence of traceability of the criteria path. Furthermore, the restriction of the time period with the operator between led to critical usability situations because this operator does not implicitly process the time specification when only 1 date is entered for a before or after query. The missing display of the codes when the selected criteria appear in the search interface also resulted in ambiguity.

The usability catastrophes and major usability problems are visualized in Multimedia Appendix 4, and the associated optimizations are suggested.

**Ontology-Specific Aspects**

The study participants assessed the orientation at the upper level of the category tree as good. In addition, it was observed that the orientation at lower levels was perceived as comprehensible by the test participants if they had background knowledge about the criteria. Overall, most of the participants (14/22, 64%) found it quite easy to identify relevant criteria as shown in Figure 7 (item 2). However, it was often observed that the display of identical or similar criteria in the free-text results list led to uncertainty in identifying relevant criteria. This was partly because of the lack of a path display, as described in the previous subsection, and partly because of the complexity and ambiguity of the ontology (eg, criteria such as glucose, glucose/BK, and glucose/blood have identical paths).

The mixed use of German and English terms—predetermined by the MII core data set—was perceived as cumbersome by some test participants and led to comprehension problems. The sorting of the criteria in the category tree was criticized at several points, and preferred alternatives were suggested; for example, some of the participants (4/22, 18%) wanted the criteria to be ordered alphabetically, whereas others (2/22, 9%) preferred sorting by relevance. Furthermore, criteria with the designation Other (...) were expected to be placed at the end of the list. When searching for female patients, it was not clearly apparent that sex had to be selected to add the characteristic female. Test participants expected the characteristic female to be selected directly in the category tree. Furthermore, some of the test participants (3/22, 14%) found the category tree to be textually overloaded.

**Feedback on Additional Features**

With regard to the additional features presented in the supplementary mock-ups, the interview analysis revealed that the implementation of the group function was considered successful and intuitive by almost all of the test participants (21/22, 95%). However, it was also pointed out that the NEW GROUP button should be made clearer and more obvious and that the assignment of characteristics to the respective desired groups should be made as simple as possible, requiring only a few clicks.

The option to link subgroups within a group in terms of temporal dependencies was perceived as rather complex. In principle, the function is considered useful because questions with temporal dependencies occur frequently, especially in the oncology field. However, the presented implementation of the function was still perceived as not very intuitive. Possible improvements could involve providing (1) a stronger emphasis of the button TIME LINKAGE, (2) context-specific information via mouse-over text, (3) a link to a brief How to section, and (4) a tutorial explaining this feature.

The discussion on the depth of criteria nesting provided a heterogeneous picture. Regarding the intuitive approach, the recorded solutions varied from the entry of individual criteria and the formation of groups to the desired possibility of assigning criteria directly to other related criteria (eg, assigning the criterion subcutaneous to the criterion insulin). The majority of the participants (14/22, 64%) would have solved the example task via groups, but this can only serve as a rough orientation because the task in the form set could only be solved theoretically and could not be worked out using the tool.

**Discussion**

**Overview**

The rationale for this work was to simultaneously develop and assess the feasibility tool v2 regarding usability and to evaluate the comprehensibility of the underlying ontology with regard to the findability of criteria.

**Discussion of Methods**

Thinking aloud tests are an established method for formative evaluations to identify usability problems and their causes early in the development process and have been applied several times in the clinical field for usability evaluation of query builders [3,8,19]. However, because of their qualitative nature, thinking aloud tests do not allow a quantitative evaluation of usability. This methodological disadvantage was compensated for by using the SUS to obtain an overall statement about how well the design of the feasibility tool v2 has succeeded. The SUS is a standardized instrument that can be used for any type of system, and it can provide valid insights into whether and to what extent usability problems exist [20]. The SUS has also been used in clinical settings for query builders [19,21].

In addition, we conducted user interviews, which are fundamentally well suited to elicit user desires and insights and have been applied several times for usability evaluations [22,23]. As our goal was not to perform statistical analyses but to collect preferences and suggestions for improvement, this method was an adequate choice.

Usability tests were conducted with a sample of 22 participants. This number is sufficient from the point of view of conducting (1) the thinking aloud test, which requires a minimum of 3 to 5 test participants [24]; (2) the SUS, which requires approximately 12 persons to reach an apparent asymptote [25];
and (3) user interviews, which require approximately 12 persons for researchers to obtain sufficient information about user problems [26].

Overall, the combination of methods allowed us to obtain a very diverse picture of user views and identify important usability issues that would need to be addressed in the next iteration. Furthermore, this combination of methods was easy to apply without the need for any special application knowledge and could be performed within a reasonable amount of time to obtain ideas for further developments very promptly.

Discussion of Results

The evaluation of the usability of the feasibility tool v2 indicated a good degree of user-friendliness. The quantitative evaluation of the SUS questionnaire also confirmed the impression gathered through user feedback. In comparison with the previous version of the prototype, it can be stated that the critical usability problems identified in the evaluation by Sedlmayr et al [8], such as the difficulty in distinguishing between inclusion and exclusion criteria or the unclear linkage using the Boolean operators, could be successfully solved and occur only in negligible numbers. There were 5 usability catastrophes in the feasibility tool v2 and 8 major problems; in comparison, there were 8 usability catastrophes and 4 major problems in the previous iteration. In this respect, there were individual improvements in usability; however, overall, there is still a need for adjustments. No comparisons can be made regarding the SUS score because the previous version was evaluated using a different set of methods (user interviews instead of web-based questionnaires), which is not unusual in iterative user-centered development [27].

Comparing the feasibility tool v2 with similar tools, it can be stated that it performs relatively well. With a SUS score of 81.15, the feasibility tool v2 performed better than the query tools Informatics for Integrating Biology and the Bedside (i2b2) [28] (SUS score=59.83); ATLAS, developed by Observational Health Data Sciences and Informatics (OHDSI) [29] (SUS score=72.81); and the GBA Sample Locator [30] (SUS score=77.03), whose user-friendliness was examined in a study published in 2021 [3]. In addition to these positive results, it must be mentioned that there is still potential for improvement. Along with making further adjustments in the area of design and refinements in navigation, the focus now is on the new features and underlying ontology.

Although the group function is technical and graphical rather easy to implement, the temporal link is more complex. Methods for technical as well as graphical implementation can already be found in the literature [31]; for example, search tools such as the aforementioned i2b2 and ATLAS take a text-based approach to display, and there are also graphical solutions such as QueryMarvel [32]. Challenges in this regard arise primarily in the technical implementation as well as in a matching intuitive presentation that should enable error-free use. The aforementioned approaches, in conjunction with the feedback from the evaluation study, will play a vital role in the deliberations that will be conducted for the next iteration process.

We also discovered that the underlying ontology has a crucial impact on the usability and acceptance of a feasibility tool. This was particularly evident in the direct comparison between the extended version with the comprehensive MII core data set [10] and the previous version with the rather lean GECCO [9]. Although there were hardly any difficulties in selecting the criteria searched for in the feasibility tool v1 [8], it was observed in the feasibility tool v2 that the search required extra time because of the more extensive ontology. It should be noted that navigation through the category tree as well as via the free-text search depends on the existing background expertise of the user. Participants with knowledge of medical terminology found it easier to navigate the category tree, whereas participants who were not familiar with relevant classifications, such as ICD-10 codes or Logical Observation Identifiers Names and Codes (LOINC) codes, had to resort to the trial-and-error method at times. This observation is also reflected in the results of the SUS score evaluation by the professionals; for example, study participants who had a medical background rated the tool better (SUS score=78.13) than those who had a biobanking background (SUS score=70.00) and tend to come from a natural science background and are unfamiliar with diagnostic and laboratory codes. This is also in line with the findings from the study comparing the 3 feasibility platforms [3], which strongly suggest that tools with more functionalities and a more extensive ontology have a harder time providing an intuitive interface. This confirms the appropriateness of our approach, which involves conducting regular evaluations based on the user-centered design process [33], thus enabling us to directly incorporate user feedback into further iterations.

Limitations

Despite the efforts we made to apply a real-world approach to the study design to obtain meaningful results for the subsequent development steps, our work includes some limitations. First, it should be mentioned that test participants were recruited for the study at sites that were ABIDE project partners. Nevertheless, care was taken to ensure that the participants were not directly involved in the project work so that they could provide an unbiased evaluation. Another aspect that could have contributed to selection bias is the fact that the study was conducted via Zoom. This method saves time and resources, but it lacks the advantages offered by a standardized test environment, although the literature shows that remote testing can be expected to produce results similar to those of laboratory testing and is an equally good method for usability testing [34,35]. As our study was conducted remotely, it is possible that mainly people with basic IT skills signed up to participate. In fact, all participants indicated that they had at least an intermediate understanding of IT. Thus, we cannot eliminate the possibility that we lack input from people who have no or little general IT expertise. Nevertheless, it can be assumed that this group of people will not be among the main users of the ABIDE feasibility tool.

Another limitation is that a prototype was evaluated. On the one hand, this had the consequence that neither test data nor real data were connected; thus, no realistic results could be provided after the query was sent. As this is only a small aspect, and the focus was on the general usability of the tool, it can be
assumed that this factor is negligible. On the other hand, because the prototype did not contain all functionalities, the envisaged additional functions could only be presented in the form of mock-ups. In this way, the analysis of the navigation path and usability was limited. However, because the evaluation took place during development, we see it as an advantage that the planned implementation could first be tested using the mock-ups before any programming work was done. According to the feedback, a revised version of the mock-ups can now be created, which can then serve as a reference for the development team.

We would like to point out that, under certain circumstances, the different ways of presenting scenarios (2 tasks in tabular format and 1 in free-text format), test execution time, and the current fatigue state of the participants could have had a possible influence on the results. However, we conducted an exploratory study with a focus on collecting suggestions for improvement for the next iteration and not a classical experiment where it is common to perform a confounder analysis.

The exclusive use of the SUS as a standardized questionnaire could be perceived as an additional limitation. Although the SUS has been used previously to evaluate ontologies, it had to be adapted for this purpose [36]. Consequently, a scale for assessing the usability of ontologies—the Ontology Usability Scale [37]—was developed, which adapts the SUS items and tailors them to ontologies. We have refrained from such a detailed evaluation of the ontology and limited ourselves to 4 items. We specifically focused on usability in the sense of ease of use, meaning that an extended consideration of the ontology would have exceeded the time frame of our study. Moreover, and this is probably the more essential point, we have no immediate influence on the ontology because it is a direct representation of the terminology used in the MII core data set, and because this is the responsibility of other working groups outside the ABIDE project, we cannot optimize it independently based on the results. Nevertheless, it was our intention not to completely disregard the ontology to identify usability problems that occur because of the underlying ontology. If these cannot be compensated by changes in the graphical user interface, we will forward the documented problem areas as the basis for discussion to the responsible persons.

Conclusions

The findings from the evaluation indicate that the investigated prototype of the feasibility tool v2 has good usability. The global SUS score of 81.25 can be rated as good. The collected feedback supports this result because the tool was frequently described as intuitive and user friendly. However, the analysis of user feedback also revealed areas that need revision. For our next development iteration, for example, we see potential for optimization above all in the display of the search functions, the unambiguous distinguishability of criteria and visibility of their associated classification system, and the implementation of the temporal linking of criteria for which recommendations for improvement will be developed. Furthermore, the findings on the comprehensibility of the ontology will be fed back to the responsible departments so that corrections can be made here as well. Overall, it can be stated that the combination of different tools used to evaluate the feasibility tool v2 provided a comprehensive view of its usability. As a development-accompanying method, we can recommend this in the planning and implementation of similar projects to be able to closely control the course of development and correct it if necessary.

Acknowledgments

The authors would like to thank the staff all participating Aligning Biobanking and Data Integration Centers Efficiently (ABIDE) locations: Aachen, Augsburg, Berlin, Dresden, Erlangen, Frankfurt, Freiburg, Gießen, Göttingen, Halle, Heidelberg, Jena, Lübeck, and Tübingen. A very special thanks goes to all scientists and researchers who participated in this study and gave us a valuable insight into the usability of the feasibility tool v2 through their loud thoughts and interview responses. The authors would also like to thank Björn Kroll, who was significantly involved in the development of the first version of the prototype, and Caroline Glathe, who assisted with taking notes during the evaluation and provided her feedback on the mock-ups. The study was conducted as part of the ABIDE project, which is funded by the German Federal Ministry of Education and Research (FKZ 01ZZ2061A, 01ZZ2061I, 01ZZ2061E, and 01ZZ2061D).

Authors’ Contributions

CS wrote the first version of the manuscript. MZ and CS planned and conducted the usability study, which was supervised by HUP and BS. JG was the team leader. TK played a leading role in the development of the graphical user interface, and LR led the work on the ontology. CS and HUP were responsible for the recruitment of the test participants. MZ and CS analyzed all thinking aloud and interview protocols as well as the recorded screen videos. All authors read the first version of the manuscript and provided valuable suggestions for changes.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Test tasks of the usability evaluation of the Aligning Biobanking and Data Integration Centers Efficiently feasibility tool v2. [PDF File (Adobe PDF File), 71 KB - humanfactors_v10i1e43782_app1.pdf]
Multimedia Appendix 2
Questionnaires of the usability evaluation of the Aligning Biobanking and Data Integration Centers Efficiently feasibility tool v2.

[PDF File (Adobe PDF File), 138 KB - humanfactors_v10i1e43782_app2.pdf]

Multimedia Appendix 3
Scoring system for the assessment of the correctness of task processing.

[PDF File (Adobe PDF File), 116 KB - humanfactors_v10i1e43782_app3.pdf]

Multimedia Appendix 4
Usability catastrophes and major usability problems (visualization and optimization).

[PDF File (Adobe PDF File), 766 KB - humanfactors_v10i1e43782_app4.pdf]

References


33. Göttgens I, Oertelt-Prigione S. The application of human-centered design approaches in health research and innovation: a narrative review of current practices. JMIR Mhealth Uhealth 2021 Dec 06;9(12):e28102 [FREE Full text] [doi: 10.2196/19650]


Abbreviations

- **ABIDE**: Aligning Biobanking and Data Integration Centers Efficiently
- **CODEX**: COVID-19 Data Exchange Platform
- **DIC**: data integration center
- **FDPG**: Research Data Portal for Health (Forschungsdatenplattform für Gesundheit)
- **FHIR**: Fast Healthcare Interoperability Resources
- **GBA**: German Biobank Alliance
- **GBN**: German Biobank Node
- **GECCO**: German Corona Consensus Data Set
- **i2b2**: Informatics for Integrating Biology and the Bedside
- **ICD-10**: International Classification of Diseases, Tenth Revision

https://humanfactors.jmir.org/2023/1/e43782

JMIR Hum Factors 2023 | vol. 10 | e43782 | p.753

(page number not for citation purposes)
Attitudes Toward the Adoption of 2 Artificial Intelligence–Enabled Mental Health Tools Among Prospective Psychotherapists: Cross-sectional Study

Anne-Kathrin Kleine¹, PhD; Eesha Kokje¹, PhD; Eva Lermer¹,², PhD; Susanne Gaube¹, PhD

¹Department of Psychology, Ludwig Maximilian University of Munich, Munich, Germany
²Technical University of Applied Sciences Augsburg, Augsburg, Germany

Corresponding Author:
Anne-Kathrin Kleine, PhD
Department of Psychology
Ludwig Maximilian University of Munich
Geschwister-Scholl-Platz 1
Munich, 80539
Germany
Phone: 49 1709076034
Email: Anne-Kathrin.Kleine@psy.lmu.de

Abstract

Background: Despite growing efforts to develop user-friendly artificial intelligence (AI) applications for clinical care, their adoption remains limited because of the barriers at individual, organizational, and system levels. There is limited research on the intention to use AI systems in mental health care.

Objective: This study aimed to address this gap by examining the predictors of psychology students’ and early practitioners’ intention to use 2 specific AI-enabled mental health tools based on the Unified Theory of Acceptance and Use of Technology.

Methods: This cross-sectional study included 206 psychology students and psychotherapists in training to examine the predictors of their intention to use 2 AI-enabled mental health care tools. The first tool provides feedback to the psychotherapist on their adherence to motivational interviewing techniques. The second tool uses patient voice samples to derive mood scores that the therapists may use for treatment decisions. Participants were presented with graphic depictions of the tools’ functioning mechanisms before measuring the variables of the extended Unified Theory of Acceptance and Use of Technology. In total, 2 structural equation models (1 for each tool) were specified, which included direct and mediated paths for predicting tool use intentions.

Results: Perceived usefulness and social influence had a positive effect on the intention to use the feedback tool (P<.001) and the treatment recommendation tool (perceived usefulness, P=.01 and social influence, P<.001). However, trust was unrelated to use intentions for both the tools. Moreover, perceived ease of use was unrelated (feedback tool) and even negatively related (treatment recommendation tool) to use intentions when considering all predictors (P=.004). In addition, a positive relationship between cognitive technology readiness (P=.02) and the intention to use the feedback tool and a negative relationship between AI anxiety and the intention to use the feedback tool (P=.001) and the treatment recommendation tool (P<.001) were observed.

Conclusions: The results shed light on the general and tool-dependent drivers of AI technology adoption in mental health care. Future research may explore the technological and user group characteristics that influence the adoption of AI-enabled tools in mental health care.

(JMIR Hum Factors 2023;10:e46859) doi:10.2196/46859

KEYWORDS
artificial intelligence; mental health; clinical decision support systems; Unified Theory of Acceptance and Use of Technology; technology acceptance model
Introduction

Background

In spite of the growing efforts to create user-friendly artificial intelligence (AI) applications, their use in clinical care remains limited [1]. Barriers to the adoption of AI-enabled clinical decision support systems (AI-CDSSs) can be found at the individual (eg, end user’s lack of trust in the system), organizational (eg, capacity to innovate), and system (eg, political decisions) levels [2-4]. Often, the adoption of AI-CDSSs fails because system and organizational requirements are not met, and accordingly, tools do not become available to potential end users [5]. The lack of regulatory oversight and standardization of AI-CDSSs can create uncertainty in the field, potentially leading to liability issues at the organizational and system levels [5]. If the system and corporate requirements for implementing a given technology are satisfied, their successful deployment depends on the practitioner’s willingness to use them. However, clinicians may be skeptical about using AI-CDSSs because of concerns regarding the accuracy and reliability of AI-generated decisions. Several frameworks and theories have been developed to systematically study the mechanisms influencing the implementation of technology in practice [5-9]. The 2 most relevant models for individual-level predictors are the Technology Acceptance Model (TAM) [10] and the Unified Theory of Acceptance and Use of Technology (UTAUT) [11]. The TAM aims to explain why a given technology is rejected or accepted by the end user. It proposes that system use is centrally driven by its perceived usefulness and ease of use. Both beliefs are determinants of attitudes toward use, which, in turn, influence use behavior [10]. The UTAUT combines the principles of 8 technology acceptance models, including the TAM. In addition to perceived usefulness (ie, performance expectancy) and perceived ease of use (ie, effort expectancy), it considers social processes (ie, social influence) and demographic variables (ie, age and gender) as predictors of use intention [11]. Accordingly, we focused on the UTAUT as the most holistic use prediction model.

Several studies have already demonstrated the applicability of the UTAUT in investigating the implementation of AI-CDSSs [12-17]. However, only 1 study has examined the predictors of the intention to use AI-enabled tools in mental health care [17]. The authors asked psychology students about their general knowledge of and attitudes toward AI systems. The results suggest a link between the perceived social norms, perceived ease of use, perceived usefulness, and perceived knowledge with students’ intention to use AI-enabled tools. However, prospective and current mental health practitioners may have varying levels of skepticism about implementing AI technology for different purposes in their (future) practice. For example, when presented with AI-generated feedback regarding diagnostic or treatment decisions, they may be reluctant to accept AI-based recommendations because of the far-reaching consequences of erroneous predictions or because they feel undermined in their role as therapists. At the same time, they may be open to incorporating AI-generated feedback regarding their interviewing techniques. Although research has begun to examine practitioners’ acceptance of AI-enabled tools in mental health care, there is a lack of specificity in assessing use intention, limiting the utility of these findings in informing practice. This study sought to address this gap by examining the intention to use two specific AI-enabled mental health tools: (1) a psychotherapy feedback tool (FB tool) that analyzes data from therapist-patient conversations and provides performance-specific feedback to the therapist [18-21] and (2) a treatment recommendation tool (TR tool) that uses voice recordings and mood scores to generate recommendations for psychotherapeutic support [22]. The research model is shown in Figure 1.
Figure 1. The research model is without control variables. The model is adapted from the preuse part of the model presented in the study by Venkatesh et al [23]. In this study, we extended the original model by adding tool understanding and cognitive technology readiness as predictors of perceived usefulness, perceived ease of use, and trust. AI: artificial intelligence.

The AI-Enabled FB Tool

Providing supervision and performance feedback during and after psychotherapy sessions enhances trainees’ and therapists’ skills acquisition and retention [20, 23]. However, these processes are labor and cost intensive and thus rarely used in training and clinical practice. Often, feedback is based on trainees’ self-reports and is only available long after the therapy session has concluded [20]. AI technology may help to reduce this problem by providing continuous, immediate, and performance-specific feedback to psychotherapists and trainees. Over the past few years, several AI-enabled FB tools have been developed and are already used in practice [24]. For example, the Therapy Insights Model uses real-time chat messages exchanged between therapists and patients to provide feedback on topics covered in the session and generate recommendations regarding topics that should be addressed in the following session [18]. Counselor Observer Ratings Expert for Motivational Interviewing uses audio recordings of motivational interviewing (MI) sessions to generate feedback on psychotherapists’ adherence to MI principles. The generated feedback focuses on 6 aspects of MI fidelity: empathy, MI spirit, reflection-to-question ratio, percent open questions, percent complex reflections, and percent MI adherence [19]. The tool chosen for this study was developed based on the Counselor Observer Ratings Expert for Motivational Interviewing. Participants were presented with information on how speech data recorded during a psychotherapy session were processed and analyzed using machine learning models to generate feedback for psychotherapists regarding their adherence to MI principles and possibilities for improvement, as shown in Figure 2.
The AI-Enabled TR Tool

Timely psychotherapeutic support may lower the risk of worsening depressive symptoms and suicidality [25]. Multiple studies have demonstrated the effectiveness of AI-enabled emotion analysis in assessing patients’ depressive states and recommending timely intervention, thereby improving mental health care [22,26]. In particular, systems have been developed in recent years to monitor or evaluate the mood of individuals with mental disorders, such as major depressive or bipolar disorder, using speech data [27,28]. These tools usually require patients to record voice samples on their mobile phones, which are analyzed by an automated speech data classifier to assess their current mood [27]. Mental health practitioners can then use this information to decide whether urgent intervention is needed [29]. The TR tool chosen for this study was based on the system developed by SondeHealth [30]. Specifically, participants were presented with information on how voice data recorded on a mobile device are processed and analyzed to generate a mood score that may be used for treatment-related decisions, as shown in Figure 3.
Research Model and Hypotheses

The first goal of this study was to test the applicability of a modified version of the UTAUT in the mental health context to understand the factors that influence the intention to use 2 specific AI-enabled mental health care tools [11,17,31,32]. In line with the UTAUT, we propose tool-specific perceived usefulness (ie, the degree to which an individual believes that using a system will enhance their performance) and perceived ease of use (ie, the degree of ease associated with using the technology) to predict the behavioral intention to use the tools in their future work. The hypotheses for this research have been preregistered through the Open Science Framework [33]. We propose the following hypotheses:

- **Hypothesis 1:** There is a positive relationship between perceived usefulness and the intention to use the tools in psychotherapy.
- **Hypothesis 2:** There is a positive relationship between perceived ease of use and the intention to use the tools in psychotherapy.

Unlike experienced psychotherapists, psychology students and psychotherapists in training may be less likely to be influenced by established work habits or procedures, which could impede the adoption of new AI technologies [11]. However, it has been suggested that students are more likely to be affected by their peers and the values and standards of their potential future employers [34]. As a result, we propose that the UTAUT variable, “social influence” (ie, the perception that other significant people think the system should be used), should be considered a predictor of students’ intention to use the tools.

- **Hypothesis 3:** There is a positive relationship between social influence and the intention to use the tools in psychotherapy.

It has been suggested that trust may be a relevant predictor of the intention to use a technology if the risk associated with it is high [12]. Because of the sensitive nature of the recommendations made by the 2 tools, we hypothesized that trust may be a predictor of students’ intention to use the tools.

- **Hypothesis 4:** There is a positive relationship between trust in the tools and the intention to use them in psychotherapy.

A lack of understanding of the underlying mechanisms of AI-enabled tools in mental health care has led to skepticism regarding their use [35,36]. In particular, the lack of transparency and explainability of AI-based clinical decision-making has impeded the adoption of such tools in mental health care [35-37]. Building on the new framework for theorizing and evaluating Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies [2], we proposed that knowledge regarding technology is a predictor of its perceived value. Consequently, we suggested that students with the knowledge and skills to apply the tools and understand how the recommendations are derived are more likely to perceive them as useful [17,38]. To test this, we extended the UTAUT by including cognitive technology readiness as an indicator of general AI knowledge and understanding of the tool as an indicator of specific AI knowledge as predictors of perceived usefulness, perceived ease of use, and trust. We preregistered 2 research questions to test this relationship:

- **Research question 1:** Is the positive relationship between cognitive technology readiness and the intention to use the tools mediated through (1) perceived usefulness, (2) perceived ease of use, and (3) trust in the tools?
- **Research question 2:** Is the positive relationship between understanding of the tools and the intention to use the tools...
mediated through (1) perceived usefulness, (2) perceived ease of use, and (3) trust in the tools?

**Methods**

**Participants**

Psychology students and psychotherapists in training were recruited through social media postings, email correspondence with administrative offices of universities, and psychotherapy training centers, as well as through the professional research-focused panel company, Prolific. Data were collected between October 2022 and January 2023, resulting in a total of 362 participants beginning the questionnaire. Of these, 208 provided answers on the behavioral intention to use the tools, resulting in a 42.54% dropout rate. In addition, 2 participants failed at least 2 of the 4 attention check items [39], leaving us with a final sample size of 206.

The final sample consisted of 16% (33/206) of men, 80.1% (165/206) of women, and 3.8% (8/206) of nonbinary individuals. The age of the participants ranged from 18 to 54 (mean 28.10, SD 7.03) years. Data were collected from Germany, the United States, the United Kingdom, and Canada. Most participants studied in Germany (111/206, 53.9%), followed by the United Kingdom (49/206, 23.8%), the United States (32/206, 15.5%), Canada (13/206, 6.3%), and other countries (1/206, 0.5%).

Regarding the field of study, most participants stated that their respective tool were assessed. Finally, we asked them about the understanding of the tool, and the intention to use the usefulness, perceived ease of use, social influence, and trust), presentation, the UTAUT predictor variables (ie, perceived psychotherapy practice/training. After the presentation, you will be asked a couple of questions about the tool.” After each tool explained how recommendations for the AI-enabled TR tool and FB tool were generated (the material is available from the first author upon request). Before seeing the slides, participants read the following short introduction: “On the following page, you will be presented with a tool that is used to [FB tool: provide feedback to psychotherapists about what went well and what could be improved in their sessions; TR tool: generate a mood score to rate the severity of patients’ depression. The mood score may be used by psychotherapists to decide which patient to treat first if multiple patients seek treatment and there is limited capacity]. Please read the information carefully and try to understand what the tool does and how it may be used in psychotherapy practice/training. After the presentation, you will be asked a couple of questions about the tool.” After each tool presentation, the UTAUT predictor variables (ie, perceived usefulness, perceived ease of use, social influence, and trust), the understanding of the tool, and the intention to use the respective tool were assessed. Finally, we asked them about their demographic information.

**Ethics Approval**

The Institutional Review Board of the University of Regensburg approved the study protocol (22-3096-101).

**Measurement Instruments**

**Independent Variables**

We assessed cognitive technology readiness with 5 items of the cognition factor of the medical AI readiness scale [40]. This scale measures terminological knowledge about medical AI applications. In total, 2 items with factor loadings<0.40 [41] that did not relate to a general understanding of AI (ie, “I can define the basic concepts of data science” and “I can define the basic concepts of statistics”) were removed. We retained 3 items related to AI understanding (ie, “I can explain how AI systems are trained,” “I can define the basic concepts and terminology of AI,” and “I can properly analyze the data obtained by AI in healthcare”; \(\alpha=.77; \omega=0.75\)).

Perceived usefulness, perceived ease of use, and social influence were measured using items adapted from the study by Venkatesh et al [32]. Participants rated their agreement on a 5-point Likert scale ranging from 1=strongly disagree to 5=strongly agree. Perceived usefulness was assessed using 5 items (eg, “Using the AI tool would enable me to accomplish tasks more quickly”). The reliabilities are \(\alpha_{FB tool}=0.86; \alpha_{TR tool}=0.91\) for the first tool and \(\alpha_{TR tool}=0.91; \alpha_{TR tool}=0.93\) for the second tool. Perceived ease of use was measured using 4 items (eg, “My interaction with the AI tool will be clear and understandable”; \(\alpha_{FB tool}=0.84; \alpha_{FB tool}=0.89; \alpha_{TR tool}=0.89; \alpha_{TR tool}=0.93\). Social influence was measured with 5 items (eg, “In my future job as a psychotherapist, people who are important to me will think that I should use the AI tool”; \(\alpha_{FB tool}=0.88; \alpha_{FB tool}=0.94; \alpha_{TR tool}=0.91; \alpha_{TR tool}=0.95\). Trust was measured with 3 items adapted from the study by Venkatesh et al [42] (eg, “The AI tool will provide access to sincere and genuine feedback”; \(\alpha_{FB tool}=0.83; \alpha_{FB tool}=0.84; \alpha_{TR tool}=0.89; \alpha_{TR tool}=0.89\). Finally, understanding of the AI-enabled tools was assessed with a single item (“Please rate your understanding of the AI-enabled feedback tool”), with answers ranging from 1=I don’t understand the tool at all to 6=I understand the tool extremely well.

**The Behavioral Intention to Use the Tools as the Dependent Variable**

The behavioral intention to use the tools was measured on a 5-point Likert scale, ranging from 1=strongly disagree to 5=strongly agree, with 3 items adapted from the study by Venkatesh et al [32] (eg, “I intend to use the AI tool in my future job as a psychotherapist”; \(\alpha_{FB tool}=0.95; \alpha_{FB tool}=0.95; \alpha_{TR tool}=0.96; \alpha_{TR tool}=0.96\).

**Control Variables**

Data privacy concerns and AI anxiety (ie, fears and insecurity regarding AI technology) have repeatedly been identified as negative predictors of the intention to use AI technology [43]. In addition, it has been shown that male participants have more positive attitudes toward AI technologies than female participants [44]. Finally, some evidence exists for the association of AI acceptance with age [45] and country [46]. Accordingly, data privacy and security concerns [47] (\(\alpha_{FB tool}=0.84; \alpha_{FB tool}=0.85; \alpha_{TR tool}=0.89; \alpha_{TR tool}=0.91\); eg, “I would be concerned that the AI tool would share my personal
information with third-parties”), AI anxiety [32] \((\alpha_{FB}=0.78; \alpha_{TR}=0.76; \omega_{FB}=0.81; \omega_{TR}=0.79)\); eg, “I feel apprehensive about using the AI tool”), gender \((0=\text{man} \text{ and } 1=\text{woman and nonbinary})\), age, and study country \((1=\text{Germany} \text{ and } 0=\text{English-speaking countries})\) were included as control variables. One item of the AI anxiety scale and 3 items of the data privacy scales with standardized factor loadings <0.40 were excluded [41].

**Data Analysis**

Data were analyzed using R software (version 4.2.2; R Foundation for Statistical Computing) [48]. First, we calculated descriptive statistics, including mean values, SDs, and correlations between study variables for each tool. Second, a confirmatory factor analysis of perceived usefulness, perceived ease of use, social influence, trust, cognitive readiness, specific tool understanding, behavioral intention to use the tool, AI anxiety, and data privacy concerns was conducted using the lavaan package [49]. We assumed at least reasonable fit for models with comparative fit index (CFI) and Tucker-Lewis index (TLI) values close to or exceeding 0.90 [50]. Root mean square error of approximation (RMSEA) values <0.08 are considered acceptable [51]. Finally, standardized root mean square residual (SRMR) values up to 0.08 are considered satisfactory [50]. We compared the theoretical measurement model with 3 more parsimonious models (combining cognitive readiness and tool understanding; perceived usefulness and ease of use; and AI anxiety and data privacy concerns) to assess whether the model variables were sufficiently distinct. Third, we conducted structural equation modeling (SEM) using the lavaan package [49] to examine the relationships between the predictor variables and the intention to use the tools to answer hypotheses 1 to 4 and research questions 1 and 2. We specified 2 models (1 for each tool) with direct effects and the mediation of the relationship between specific tool understanding, cognitive AI readiness, and the intention to use the tool. We followed the recommendations by Scharf et al [52] to determine whether the regression coefficients should be regularized. Specifically, we applied regularization in case of multicollinearity and associated inflated SEs [52]. The study data and R script will be made available on the web on publication [33].

**Preregistration Statement**

The hypotheses were preregistered in the Open Science Framework [33]. Exploratory hypotheses were thus identified.

**Results**

Table 1 presents the means, SDs, and correlations. We specified the theoretical model with perceived usefulness, perceived ease of use, social influence, trust, cognitive readiness, specific tool understanding, behavioral intention to use the tool, AI anxiety, and data privacy concerns to load on separate factors. The theoretical model fitted the data adequately \((FB \text{ tool: } \chi^2_{370}=808.9, P<.001; \text{CFI}=0.89; \text{TLI}=0.87; \text{RMSEA}=0.08; \text{SRMR}=0.08 \text{ and TR tool: } \chi^2_{370}=713.41, P<.001; \text{CFI}=0.93; \text{TLI}=0.92; \text{RMSEA}=0.07; \text{SRMR}=0.06)\).

The theoretical model fit the data better than the 3 more parsimonious models (ie, cognitive readiness and specific tool understanding combined; \(FB \text{ tool: } \chi^2_{70}=50.37, P<.001 \text{ and TR tool: } \chi^2_{70}=72.68, P<.001\); perceived usefulness and perceived ease of use combined, \(FB \text{ tool: } \chi^2_{70}=257.79, P<.001 \text{ and TR tool: } \chi^2_{70}=435.43, P<.001\); and AI anxiety and data privacy concerns combined, \(FB \text{ tool: } \chi^2_{8}=240.91, P<.001 \text{ and TR tool: } \chi^2_{8}=133.6, P<.001\)). Thus, we concluded that the model variables were sufficiently distinct.

To test hypotheses 1 to 4 and research questions 1 and 2, we specified 2 SEMs (1 for each tool) with the behavioral intention to use \(FB \text{ tool} \) and \(TR \text{ tool} \) to be predicted by the respective UTAUT variables (ie, perceived usefulness, perceived ease of use, social influence, and trust); tool understanding; cognitive readiness; and the control variables AI anxiety, data privacy concerns, age, male gender \((0=\text{man} \text{ and } 1=\text{woman and nonbinary})\), and study country \((1=\text{Germany} \text{ and } 0=\text{English-speaking countries})\). In addition, we added mediated pathways of the relationship of specific tool understanding and cognitive AI readiness with the intention to use the tools through perceived usefulness, perceived ease of use, and trust in the tool. No inflated SEs were observed, and we proceeded with the interpretation of the SEM without regularization. The results are presented in Tables 2 and 3. Figure 4 shows the significant paths from the SEM path models. As can be seen in Tables 2 and 3 and Figure 4, the relevant paths differ between the 2 models. Perceived usefulness and social influence showed the expected positive relationships with the intention to use both tools, supporting hypotheses 1 and 3. However, trust was unrelated to use intention in both models, and perceived ease of use was unrelated to the intention to use the \(FB \text{ tool} \) and was negatively related to the intention to use the \(TR \text{ tool} \). Accordingly, we found no support for hypotheses 2 and 4. AI anxiety was negatively related to use intentions in both models.

Finally, the exploratory mediation analysis results suggest that the relationships of tool understanding and cognitive technology readiness with the intention to use \(FB \text{ tool} \) are not mediated through perceived usefulness, perceived ease of use, or trust. There was a negative mediation effect of the relationship between tool understanding and the intention to use the \(TR \text{ tool} \) through perceived ease of use, that is, tool understanding was positively related to perceived ease of use, which, in turn, was negatively associated with use intention.
Table 1. Means, SDs, and correlations among study variables

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. PU</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>-.05</td>
<td>-.20</td>
<td>.08</td>
<td>.05</td>
<td>-.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. PE</td>
<td>.44</td>
<td>.02</td>
<td>.38</td>
<td>.54</td>
<td>.23</td>
<td>-.26</td>
<td>-.37</td>
<td>.05</td>
<td>-.10</td>
<td>.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. SI</td>
<td>.59</td>
<td>.04</td>
<td></td>
<td>.19</td>
<td>.78</td>
<td>-.02</td>
<td>-.25</td>
<td>.13</td>
<td>.12</td>
<td>-.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. TR</td>
<td>.68</td>
<td>.50</td>
<td>.57</td>
<td>.19</td>
<td>.72</td>
<td>-.17</td>
<td>-.31</td>
<td>.08</td>
<td>.04</td>
<td>-.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. TU</td>
<td>.10</td>
<td>.04</td>
<td>-.01</td>
<td>.14</td>
<td>.15</td>
<td>-.21</td>
<td>-.19</td>
<td>.21</td>
<td>-.13</td>
<td>-.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. UT</td>
<td>.70</td>
<td>.49</td>
<td>.67</td>
<td>.66</td>
<td>.08</td>
<td></td>
<td>-.11</td>
<td>-.41</td>
<td>.11</td>
<td>.11</td>
<td>-.23</td>
<td></td>
</tr>
<tr>
<td>7. PC</td>
<td>.07</td>
<td>-.17</td>
<td>-.06</td>
<td>-.24</td>
<td>.10</td>
<td>-.18</td>
<td>.33</td>
<td>-.11</td>
<td>.19</td>
<td>-.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. ANX</td>
<td>-.08</td>
<td>-.31</td>
<td>-.11</td>
<td>-.21</td>
<td>-.22</td>
<td>-.52</td>
<td>.31</td>
<td></td>
<td>-.12</td>
<td>-.11</td>
<td>.16</td>
<td></td>
</tr>
<tr>
<td>9. CR</td>
<td>.07</td>
<td>.15</td>
<td>.14</td>
<td>.09</td>
<td>.21</td>
<td>.22</td>
<td>-.11</td>
<td>-.19</td>
<td>.01</td>
<td>-.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Age</td>
<td>.01</td>
<td>-.03</td>
<td>.12</td>
<td>-.06</td>
<td>-.21</td>
<td>.02</td>
<td>.11</td>
<td>-.03</td>
<td>.01</td>
<td></td>
<td>-.11</td>
<td></td>
</tr>
<tr>
<td>11. Gend</td>
<td>-.08</td>
<td>-.01</td>
<td>-.15</td>
<td>-.11</td>
<td>.00</td>
<td>-.12</td>
<td>-.02</td>
<td>.11</td>
<td>-.07</td>
<td>-.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Country</td>
<td>-.10</td>
<td>-.06</td>
<td>-.13</td>
<td>.01</td>
<td>-.02</td>
<td>-.03</td>
<td>-.02</td>
<td>-.10</td>
<td>-.02</td>
<td>-.22</td>
<td>.21</td>
<td></td>
</tr>
<tr>
<td>FB tool</td>
<td>3.2</td>
<td>3.7</td>
<td>2.9</td>
<td>3.4</td>
<td>4.2</td>
<td>2.9</td>
<td>4.2</td>
<td>2.7</td>
<td>2.5</td>
<td>28.1</td>
<td>0.8</td>
<td>0.5</td>
</tr>
<tr>
<td>TR tool</td>
<td>2.8</td>
<td>3.9</td>
<td>2.7</td>
<td>3.0</td>
<td>4.5</td>
<td>2.4</td>
<td>4.0</td>
<td>2.9</td>
<td>2.5</td>
<td>28.1</td>
<td>0.8</td>
<td>5.0</td>
</tr>
</tbody>
</table>

The lower triangle of the correlation table contains the correlations for the FB tool, and the upper triangle contains the correlations for the TR tool. All correlations ≥|0.14| are significant at P<.05.

bPU: perceived usefulness.
cNot applicable.
dPE: perceived ease of use.
eSI: social influence.
fTR: trust in the tool.
gTU: tool understanding.
hIU: intention to use the tool.
iPC: privacy concerns.
jANX: artificial intelligence anxiety.
kCR: cognitive technology readiness.
lCode: 0=man and 1=woman and nonbinary.
mCode: 1=Germany and 0=English-speaking country.

FB tool: feedback tool.
TR tool: treatment recommendation tool.
<table>
<thead>
<tr>
<th>Effect</th>
<th>Feedback tool</th>
<th>B (SE)</th>
<th>β (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Direct effects (DV&lt;sup&gt;a&lt;/sup&gt;=IU&lt;sup&gt;b&lt;/sup&gt;)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PU&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.63 (0.11)</td>
<td>.51 (.30 to .72)</td>
<td>.001</td>
<td></td>
</tr>
<tr>
<td>PE&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.06 (0.06)</td>
<td>.03 (−.09 to .15)</td>
<td>.59</td>
<td></td>
</tr>
<tr>
<td>SI&lt;sup&gt;e&lt;/sup&gt;</td>
<td>0.37 (0.07)</td>
<td>.32 (.19 to .46)</td>
<td>.001</td>
<td></td>
</tr>
<tr>
<td>TR&lt;sup&gt;f&lt;/sup&gt;</td>
<td>0.06 (0.12)</td>
<td>.04 (−.19 to .27)</td>
<td>.72</td>
<td></td>
</tr>
<tr>
<td>CR&lt;sup&gt;g&lt;/sup&gt;</td>
<td>0.12 (0.05)</td>
<td>.12 (.02 to .22)</td>
<td>.02</td>
<td></td>
</tr>
<tr>
<td>TU&lt;sup&gt;h&lt;/sup&gt;</td>
<td>−0.07 (0.05)</td>
<td>−.07 (−.18 to .03)</td>
<td>.16</td>
<td></td>
</tr>
<tr>
<td>PC&lt;sup&gt;i&lt;/sup&gt;</td>
<td>−0.03 (0.05)</td>
<td>−.04 (−.13 to .06)</td>
<td>.42</td>
<td></td>
</tr>
<tr>
<td>ANX&lt;sup&gt;j&lt;/sup&gt;</td>
<td>−0.18 (0.06)</td>
<td>−.18 (−.29 to −.07)</td>
<td>.001</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.00 (0.04)</td>
<td>−.01 (−.10 to .07)</td>
<td>.74</td>
<td></td>
</tr>
<tr>
<td>Gender&lt;sup&gt;k&lt;/sup&gt;</td>
<td>−0.08 (0.04)</td>
<td>−.03 (−.12 to .05)</td>
<td>.48</td>
<td></td>
</tr>
<tr>
<td>Country&lt;sup&gt;l&lt;/sup&gt;</td>
<td>0.04 (0.04)</td>
<td>.02 (−.07 to .11)</td>
<td>.66</td>
<td></td>
</tr>
<tr>
<td><strong>Direct effects (DV&lt;sub&gt;s&lt;/sub&gt;=PU, PE, and TR)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TU→PU</td>
<td>0.09 (0.08)</td>
<td>.12 (.03 to .27)</td>
<td>.13</td>
<td></td>
</tr>
<tr>
<td>CR→PU</td>
<td>0.04 (0.08)</td>
<td>.04 (.12 to .20)</td>
<td>.60</td>
<td></td>
</tr>
<tr>
<td>TU→PE</td>
<td>0.24 (0.06)</td>
<td>.45 (.32 to .57)</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>CR→PE</td>
<td>0.02 (0.07)</td>
<td>.04 (.11 to .18)</td>
<td>.62</td>
<td></td>
</tr>
<tr>
<td>TU→TR</td>
<td>0.09 (0.08)</td>
<td>.13 (.02 to .28)</td>
<td>.09</td>
<td></td>
</tr>
<tr>
<td>CR→TR</td>
<td>0.07 (0.08)</td>
<td>.10 (.06 to .26)</td>
<td>.22</td>
<td></td>
</tr>
<tr>
<td><strong>Indirect effects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TU→PU→IU</td>
<td>0.06 (0.04)</td>
<td>.06 (.02 to .14)</td>
<td>.16</td>
<td></td>
</tr>
<tr>
<td>TU→PE→IU</td>
<td>0.01 (0.03)</td>
<td>.01 (.04 to .07)</td>
<td>.59</td>
<td></td>
</tr>
<tr>
<td>TU→TR→IU</td>
<td>0.01 (0.02)</td>
<td>.01 (.02 to .04)</td>
<td>.73</td>
<td></td>
</tr>
<tr>
<td>CR→PU→IU</td>
<td>0.02 (0.04)</td>
<td>.02 (.06 to .10)</td>
<td>.60</td>
<td></td>
</tr>
<tr>
<td>CR→PE→IU</td>
<td>0.00 (0.00)</td>
<td>.00 (.01 to .01)</td>
<td>.71</td>
<td></td>
</tr>
<tr>
<td>CR→TR→IU</td>
<td>0.00 (0.01)</td>
<td>.00 (.02 to .03)</td>
<td>.73</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>DV: dependent variable.  
<sup>b</sup>IU: intention to use the tool.  
<sup>c</sup>PU: perceived usefulness.  
<sup>d</sup>PE: perceived ease of use.  
<sup>e</sup>SI: social influence.  
<sup>f</sup>TR: trust in the tool.  
<sup>g</sup>CR: cognitive technology readiness.  
<sup>h</sup>TU: tool understanding.  
<sup>i</sup>PC: privacy concerns.  
<sup>j</sup>ANX: artificial intelligence anxiety.  
<sup>k</sup>Code: 0=man and 1=woman and nonbinary.  
<sup>l</sup>Code: 1=Germany and 0=English-speaking country.
Table 3. Structural equation modeling results predicting the intention to use the treatment recommendation tool.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Treatment recommendation tool</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B (SE)</td>
</tr>
<tr>
<td>Direct effects (DVa=IUb)</td>
<td></td>
</tr>
<tr>
<td>PUc</td>
<td>0.31 (0.11)</td>
</tr>
<tr>
<td>PEd</td>
<td>−0.29 (0.06)</td>
</tr>
<tr>
<td>Si</td>
<td>0.56 (0.08)</td>
</tr>
<tr>
<td>Trf</td>
<td>0.23 (0.11)</td>
</tr>
<tr>
<td>CRe</td>
<td>−0.01 (0.04)</td>
</tr>
<tr>
<td>Tuib</td>
<td>0.02 (0.05)</td>
</tr>
<tr>
<td>PCi</td>
<td>−0.01 (0.05)</td>
</tr>
<tr>
<td>Anxj</td>
<td>−0.25 (0.06)</td>
</tr>
<tr>
<td>Age</td>
<td>0.00 (0.04)</td>
</tr>
<tr>
<td>Genderk</td>
<td>−0.04 (0.04)</td>
</tr>
<tr>
<td>Countryl</td>
<td>−0.08 (0.04)</td>
</tr>
<tr>
<td>Direct effects (DVs=PU, PE, and TR)</td>
<td></td>
</tr>
<tr>
<td>TU→PU</td>
<td>0.15 (0.07)</td>
</tr>
<tr>
<td>CR→PU</td>
<td>0.04 (0.08)</td>
</tr>
<tr>
<td>TU→PE</td>
<td>0.40 (0.05)</td>
</tr>
<tr>
<td>CR→PE</td>
<td>−0.06 (0.07)</td>
</tr>
<tr>
<td>TU→TR</td>
<td>0.15 (0.07)</td>
</tr>
<tr>
<td>CR→TR</td>
<td>0.06 (0.08)</td>
</tr>
<tr>
<td>Indirect effects</td>
<td></td>
</tr>
<tr>
<td>TU→PU→IU</td>
<td>0.05 (0.03)</td>
</tr>
<tr>
<td>TU→PE→IU</td>
<td>−0.11 (0.04)</td>
</tr>
<tr>
<td>TU→TR→IU</td>
<td>0.03 (0.02)</td>
</tr>
<tr>
<td>CR→PU→IU</td>
<td>0.01 (0.02)</td>
</tr>
<tr>
<td>CR→PE→IU</td>
<td>0.02 (0.01)</td>
</tr>
<tr>
<td>CR→TR→IU</td>
<td>0.01 (0.01)</td>
</tr>
</tbody>
</table>

aDV: dependent variable.
bIU: intention to use the tool.
cPU: perceived usefulness.
dPE: perceived ease of use.
eSI: social influence.
fTR: trust in the tool.
gCR: cognitive technology readiness.
hTU: tool understanding.
iPC: privacy concerns.
jANX: artificial intelligence anxiety.
kCode: 0=man and 1=woman and nonbinary.
lCode: 1=Germany and 0=English-speaking country.
Figure 4. The results of exploratory mediation analysis. Significant paths for the prediction of the intention to use the artificial intelligence (AI)–enabled (A) feedback tool and (B) treatment recommendation tool. Only nonzero paths and indirect effects are displayed.

Discussion

Principal Findings

In recent years, there has been a rapid growth in the development of AI-enabled mental health care tools. To investigate the implementation challenges and potential user needs, in this study, we examined the intention to use 2 AI-enabled mental health care tools among psychology students and psychotherapists in training. The first tool provides feedback to the psychotherapist on their adherence to MI techniques by analyzing data collected during psychotherapy sessions. The second tool uses patient voice samples to derive mood scores that the therapists may use for treatment decisions. An extended UTAUT model was used to analyze the results, which showed that perceived usefulness and social influence had a positive effect on the intention to use both tools. However, trust was unrelated to the intention to use both tools, and perceived ease of use was unrelated (FB tool) and even negatively related (TR tool) to the intention to use when considering all predictors in 1 model.

The findings of this study are partly in line with previous research on AI-CDSSs in medicine [13,15]. Fan et al [13] found positive associations between perceived usefulness and trust with use intentions among a sample of health care professionals, and Zhai et al [15] reported positive relationships between perceived usefulness and social influence with the intention to use AI-assisted contouring technology among radiation oncologists. Furthermore, Tran et al [16] identified social influence as the only significant predictor of the intention to use AI-CDSSs among undergraduate medical students. Gado et al [17] found support for the direct effects of perceived usefulness, AI knowledge, and perceived social norms on the intention to use AI as well as indirect effects of perceived ease of use on use intention via positive attitudes toward AI in a sample of psychology students. This consistent link between social influence and AI use intentions found in studies using student samples may be explained by the greater susceptibility of students to influence of peers and prospective employers [53]. As students have yet to develop a professional identity that shapes their work-related decisions, they may be more likely to align their decisions with the perceived expectations of influential others [54].

The assessment of symptom severity often involves complex interactions with the patient and reflections on psychotherapeutic elements, which may make participants skeptical of a device that is perceived as being easy to use. One explanation for the null and negative relationships between perceived ease of use and use intentions for AI-generated recommendations in the mental health field may be the high stakes of accepting the tool’s advice. This interpretation might be supported by a study predicting intentions to learn about AI applications among medical staff [55], which found that perceived ease of use was the strongest predictor of the intention to learn how to use AI-enabled tools in health care. Combined with the results of this study, it may be assumed that ease of use positively predicts interactions with AI-generated advice that aligns with the user’s level of competency and professionalism. That is, ease of use may positively predict learning intentions but maybe not the intention to use high-stakes mental health tools among students and trainees who have not yet gained profound professional experience. Students’ primary task at university is to learn and acquire skills and knowledge. The ease with which an AI-enabled tool can be applied likely becomes more relevant when the interaction with such tools is required or advantageous for their professional performance. More research is needed to understand the conditions under which perceived ease of use is positively related to AI use intentions among medical and mental health practitioners and to explore the implications of the high stakes associated with AI-generated recommendations.

Trust in the tools was unrelated, whereas AI anxiety was negatively related to the intention to use both the FB and TR tools. One explanation for this finding may be participants’ limited insight into the functioning mechanisms of the tools. A profound assessment of their trust in the tools requires more
in-depth knowledge than assessing their AI anxiety. Specifically, whether the AI tool “will provide data in [their] best interest,” “provides access to sincere and genuine feedback,” or “will perform its role of a supportive system very well” [42] may be difficult to assess without having used the tool in practice and, thus, may be less relevant for students’ intention to use the tool. In contrast, AI anxiety represents intuitive, affective reactions, such as feeling apprehensive about the tool or being hesitant to use the tool for fear of making mistakes [32]. As students and psychotherapists in training have limited to no experience interacting with AI-generated feedback, they may base their decision-making on intuitive, emotional reactions better represented by AI anxiety than trust in the tools [56].

By differentiating between specific tool understanding and more general cognitive technology readiness, this study moves beyond previous research that focused on the role of general AI knowledge in predicting general use intention [17]. The mediation analyses revealed that none of the 3 UTAUT variables mediated the relationship between tool understanding and cognitive technology readiness with the intention to use the FB tool. However, there was a positive relationship between cognitive technology readiness and the intention to use the FB tool. This might indicate that general AI understanding may spur use intentions of low-stakes AI-generated advice but not the intention to use AI advice for deriving treatment decisions. In addition, in line with the direct effects, perceived ease of use emerged as a negative mediator between specific tool understanding and the intention to use the TR tool. The results of the exploratory mediation models highlighted the relevance of distinguishing between different AI-enabled tools when assessing the relationship between different forms of AI knowledge and use intentions.

Limitations and Future Directions

This study has some limitations. First, we collected data at only 1 time point. Although cross-sectional designs are commonly chosen to investigate mechanisms predicted by the UTAUT [13,15], they prevent the assessment of an order of effects. The adoption of AI-generated advice should be studied longitudinally to increase the understanding of use-predicting mechanisms. Second, although studying technology acceptance with deterministic models, such as the UTAUT and TAM, has a long tradition, such studies have recently been criticized for their oversimplicity, which lowers their explanatory power. In this vein, focusing on 2 specific AI-enabled mental health tools may be highlighted as a strength of this study, as it increases the ecological validity of the results. However, future research should seek to integrate organizational and system processes to provide a more profound understanding of the mechanisms that prevent and promote technology adoption. Other frameworks and theories, such as activity theory [57], adaptive structuration theory [58], and the Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies framework [2], may serve as theoretical underpinnings of research investigating use in context instead of focusing on individual-centered variables alone [5]. Finally, we focused on psychology students and psychotherapists in training as a potential user group and found discrepancies in our results compared with previous research findings [13,16]. Future research should compare adoption and adoption intentions among multiple (potential) user groups and tools to shed light on tool-dependent and user-dependent predicting mechanisms.

Conclusions

This study provides insights into the individual implementation challenges of AI-enabled FB and TR tools used in mental health care. The results highlight the relevance of specific UTAUT predictors as general drivers of AI technology adoption in mental health care (ie, perceived usefulness, social influence, and AI anxiety) and emphasize the need to distinguish between different AI technologies with reference to other influencing factors (ie, perceived ease of use, cognitive technology readiness, and tool understanding). Future research should explore the conditions under which perceived ease of use is positively related to AI use intentions among mental health practitioners.

Conflicts of Interest

None declared.

References


33. Student mental health AI tools. Open Source Framework. URL: https://osf.io/fqdzb [accessed 2023-06-26]


Abbreviations

AI: artificial intelligence
AI-CDSS: artificial intelligence–enabled clinical decision support system
CFI: comparative fit index
FB tool: feedback tool
MI: motivational interviewing
RMSEA: root mean square error of approximation
SEM: structural equation modeling
SRMR: standardized root mean square residual
TAM: Technology Acceptance Model
TLI: Tucker-Lewis index
TR tool: treatment recommendation tool
UTAUT: Unified Theory of Acceptance and Use of Technology

©Anne-Kathrin Kleine, Eesha Kokje, Eva Lermer, Susanne Gaube. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 12.07.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

https://humanfactors.jmir.org/2023/1/e46859
Intensive Care Unit Physicians’ Perspectives on Artificial Intelligence–Based Clinical Decision Support Tools: Preimplementation Survey Study

Siri L van der Meijden1,2,3, MSc; Anne A H de Hond2,4, MSc; Patrick J Thoral5, MD; Ewout W Steyerberg4, MSc, PhD; Ilse M J Kant2,4, MSc, PhD; Giovanni Cina6,7,8, MSc, PhD; M Sesmu Arbous1, MSc, MD, PhD

1Department of Intensive Care Medicine, Leiden University Medical Center, Leiden, Netherlands
2Clinical AI Implementation and Research Lab, Leiden University Medical Center, Leiden, Netherlands
3Healthplus.ai, Amsterdam, Netherlands
4Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, Netherlands
5Department of Intensive Care Medicine, Laboratory for Critical Care Computational Intelligence, Amsterdam Medical Data Science, Amsterdam University Medical Centers, Amsterdam, Netherlands
6Pacmed, Amsterdam, Netherlands
7Institute for Logic, Language and Computation, University of Amsterdam, Amsterdam, Netherlands
8Department of Medical Informatics, Amsterdam University Medical Center, University of Amsterdam, Amsterdam, Netherlands

Corresponding Author:
Siri L van der Meijden, MSc
Department of Intensive Care Medicine
Leiden University Medical Center
Albinusdreef 2
Leiden, 2333 ZA
Netherlands
Phone: 31 71 526 9111
Email: S.L.van_der_meijden@lumc.nl

Abstract

Background: Artificial intelligence–based clinical decision support (AI-CDS) tools have great potential to benefit intensive care unit (ICU) patients and physicians. There is a gap between the development and implementation of these tools.

Objective: We aimed to investigate physicians’ perspectives and their current decision-making behavior before implementing a discharge AI-CDS tool for predicting readmission and mortality risk after ICU discharge.

Methods: We conducted a survey of physicians involved in decision-making on discharge of patients at two Dutch academic ICUs between July and November 2021. Questions were divided into four domains: (1) physicians’ current decision-making behavior with respect to discharging ICU patients, (2) perspectives on the use of AI-CDS tools in general, (3) willingness to incorporate a discharge AI-CDS tool into daily clinical practice, and (4) preferences for using a discharge AI-CDS tool in daily workflows.

Results: Most of the 64 respondents (of 93 contacted, 69%) were familiar with AI (62/64, 97%) and had positive expectations of AI, with 55 of 64 (86%) believing that AI could support them in their work as a physician. The respondents disagreed on whether the decision to discharge a patient was complex (23/64, 36% agreed and 22/64, 34% disagreed); nonetheless, most (59/64, 92%) agreed that a discharge AI-CDS tool could be of value. Significant differences were observed between physicians from the 2 academic sites, which may be related to different levels of involvement in the development of the discharge AI-CDS tool.

Conclusions: ICU physicians showed a favorable attitude toward the integration of AI-CDS tools into the ICU setting in general, and in particular toward a tool to predict a patient’s risk of readmission and mortality within 7 days after discharge. The findings of this questionnaire will be used to improve the implementation process and training of end users.

(JMIR Hum Factors 2023;10:e39114) doi:10.2196/39114

https://humanfactors.jmir.org/2023/1/e39114
KEYWORDS
intensive care unit; hospital; discharge; artificial intelligence; AI; clinical decision support; clinical support; acceptance; decision support; decision-making; digital health; eHealth; survey; perspective; attitude; opinion; adoption; prediction; risk

Introduction
Due to the increasing availability of high-quality clinical data, the development of artificial intelligence–based clinical decision support (AI-CDS) tools to enhance personalized medicine is on the rise. AI-CDS tools make use of learning algorithms, including machine learning, which may, in specific circumstances, outperform classical statistical models when applied to large data sets for health care–related prediction tasks [1-3]. The complex nature of these artificial intelligence (AI) algorithms and their use of numerous input variables may lead to “black box” algorithms, which often leave it unclear why the algorithm output specific predictions [4,5]. In the intensive care unit (ICU), complex and high-stakes decisions are made that might benefit from data-driven decision support [6]. The ICU is the most data-rich environment in the hospital due to high-frequency monitoring, and there has been an increase in the literature on AI model development for ICU decision support [7]. However, a recent review showed that implementation of these AI-CDS tools in clinical ICU practice is lacking due to difficulties at several levels [8]. These difficulties include patient privacy, regulatory aspects, and a lack of demonstrations of these tools’ clinical value in the complex ICU environment [8]. To enhance clinical uptake and integration in daily workflows and to tailor AI-CDS tools to physicians’ needs, we need a broad understanding of physicians’ current decision-making practices and their views on the use of AI-CDS tools [9-11].

There is a need to study human factors for the safe and effective implementation of AI-CDS tools, as high predictive performance does not ensure acceptance of these technologies [12,13]. Physicians’ perspectives on clinical AI have been investigated in survey studies in the fields of psychiatry [14], gastroenterology [15], diagnostic pathology [16], and cardiology [17], as well as across specialties [18-20]. In general, strong interest and favorable attitudes toward the use of AI-CDS were reported, but no study has focused solely on the application of AI-CDS tools in the ICU in terms of willingness to use such a tool in clinical practice and how it would fit into clinical workflows. As the ICU is unique in terms of the complexity of decisions, the pressure under which decisions have to be made, and the potential in terms of data availability, knowledge focused on this clinical domain is highly relevant. To understand the potential of AI-CDS tools in the data-rich ICU environment, and to attempt to solve the challenging “last mile” problem facing real-world implementations, we need to gather more insights on clinicians’ attitudes and perspectives regarding this subject in the local context [21,22]. These insights may enhance successful implementation in this high-stakes decision-making environment, as clinician input is important throughout the implementation process to enhance successful implementation and ultimately improve patient outcomes [23].

This survey study is part of preimplementation research for Pacmed Critical [24]. Pacmed Critical is a machine learning–based AI-CDS tool that predicts a patient’s combined readmission and mortality risk within 7 days of ICU discharge to support physicians in their decisions to discharge patients to lower care wards [25,26]. The Pacmed Critical software is intended for use as a complementary tool by qualified ICU medical professionals and will be accessed on hospital premises; it will not be used on mobile devices. We aimed to investigate (1) physicians’ current decision-making behavior with respect to discharging ICU patients, (2) physicians’ perspectives on the use of AI decision support tools in general, (3) physicians’ willingness to incorporate an AI-CDS tool in daily clinical practice, and (4) physicians’ preferences for using an AI-CDS tool in their daily workflows. As knowledge of physicians’ attitudes toward the implementation of AI-CDS tools is currently lacking for the ICU domain, the overall aim of this survey was to investigate ICU physicians’ perspectives on AI-CDS tools to enhance the implementation process and to raise awareness among ICU physicians of an upcoming implementation.

Methods

Study Sample
The survey was conducted between July and December 2021 at Leiden University Medical Center (LUMC), Leiden, and Amsterdam University Medical Center (Amsterdam UMC, Vrije Universiteit Medical Center location), Amsterdam, both in the Netherlands. Both centers are academic tertiary referral hospitals. At Amsterdam UMC, 2 intensivists codeveloped Pacmed Critical, and other ICU clinicians took part in end user testing as part of the Conformité Européenne (CE) certification process of the software. The LUMC physicians were not involved in the development of the tool, and implementation was planned to start after completion of the survey.

Results were collected anonymously on paper at LUMC and by means of a web-based survey at Amsterdam UMC. All physicians working at the ICU were eligible to participate in this study, including residents, intensive care fellows, and board-certified intensivists.

Ethics Approval
The results of this research do not include any sensitive or identifiable data. We obtained ethical approval from the medical ethical committee of LUMC (ID: N21.153).

Survey Instrument
The survey instrument was developed with the expertise of 2 ICU physicians, AI and organizational researchers, a data scientist, a user experience researcher, and a Pacmed Critical product owner. The 20-question survey consisted of 13 statements, 5 multiple-choice questions, and 2 open questions. Statements were answered on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree) [27]. Multimedia Appendices 1 and 2 show the full questionnaire (in English and Dutch, respectively); Multimedia Appendix 1 also describes the rationale for each survey question. Participants did not receive additional background information on Pacmed Critical.

https://humanfactors.jmir.org/2023/1/e39114
other than that it used an AI algorithm based on patient data from electronic health records (EHRs). The survey was divided into 4 domains. These 4 domains and individual questions in the domains were chosen to obtain knowledge to optimize further development and enhance the implementation process.

**Physicians’ Current Decision-Making Behavior With Respect to Discharging ICU Patients (Q1-3, Q11-13)**

The aim of the questions in this domain was to investigate current decision-making and whether the discharge AI-CDS tool could be of benefit in terms of the complexity of the discharge decision and the predicted outcome. The first 3 statement questions investigated the complexity of the decision to discharge ICU patients and the influence of readmission risk and bed availability on this decision. The average certainty that a patient would not be readmitted after the decision to discharge was made was ranked on a scale from 1 (completely uncertain) to 10 (completely certain). We asked about patient groups for whom the decision to discharge was perceived as most challenging to determine where the AI-CDS tool could be of most value (these questions were multiple choice). We also asked about which factors were deemed most important in the process of discharging patients (open answers were solicited).

**Perspectives on the Use of AI-CDS Tools in General (Q4-8)**

Five statements covered perspectives and attitudes toward AI-CDS tools at the ICU, as the participants had no or little experience in working with these tools. These included statements on familiarity with AI, whether AI was believed to be able to replace physicians in the future, the anticipated added value and support of AI-CDS at the ICU, and whether AI-CDS tools represented the physicians’ work sufficiently to be of support.

**Willingness to Incorporate the Discharge AI-CDS Tool Into Daily Clinical Practice (Q9,10, 17-20)**

The willingness to incorporate the discharge AI-CDS in clinical practice was assessed with 5 statements on belief in the positive value of discharge decision support, the importance of having insight into the contributing factors to the prediction, the potential influence a prediction may have on discharge decision-making, willingness to consult the prediction before making the decision to discharge a patient, and the feasibility of incorporating the prediction into the physicians’ workflows. Furthermore, the physicians were asked to indicate the threshold of predicted readmission and mortality risk (on a scale of 0 to 100) above which they would not discharge a patient to the ward, and below which they would discharge a patient. The aim of this question was to study the influence of a certain predicted chance of readmission and mortality on the physician’s behavior.

**Preferences for Using a Discharge AI-CDS tool in Daily Workflows (Q14-16)**

The last domain included questions on how the AI-CDS tool for discharging ICU patients could be integrated into the current clinical workflow at the ICU; the answers were intended to be used as input in the design and implementation process of the AI-CDS tool, in order to make it part of current decision-making processes [11]. We used multiple-choice questions to determine the preferred method to access the predictions (ie, on a dashboard or integrated in EHRs), the preferred moment or moments to access the predictions, when the predictions should be updated, and the most relevant end users. Information gathered from these questions was used to understand the demands on the user interface and to optimize implementation and daily use. One or more options could be chosen for the multiple-choice questions. Lastly, respondents could leave open comments and suggestions.

**Data Analysis**

Results are given as percentages of the total number of respondents for categorical questions. Answers to numerical questions are summarized as the median (IQR). Because the participating physicians at the 2 centers differed in their involvement in the development of the tool, we performed separate analyses for LUMC and Amsterdam UMC for the questions in domain 1 (current decision-making behavior with respect to discharging ICU patients), domain 2 (attitudes and perspectives on AI-CDS), and domain 3 (willingness to incorporate a discharge AI-CDS tool into daily clinical practice). As an additional subgroup analysis, we investigated differences in the responses to the questions in domains 2 and 3 between intensivists and other physicians working at the ICU (ie, residents and fellows at the ICU). We determined significant associations for the Likert-scale statement questions with the Mann-Whitney U test. The level of significance was set at P<.05.

**Results**

**Sample Characteristics**

The survey was distributed to 40 clinicians at LUMC and 53 clinicians at Amsterdam UMC. A total of 64 of 93 (69%) of these clinicians completed the survey, including 33 of 64 (52%) at LUMC and 31 of 64 (48%) at Amsterdam UMC (Table 1). The total group had a median 2.75 (IQR 1-10) years of ICU work experience; the LUMC group had 3 (IQR 1-10.5) years and the Amsterdam UMC group 2 (IQR 1.5-10) years (P=.94). In the Netherlands, medical residents from many specialties are assigned a rotation in the ICU as part of their specialist training, which is reflected by the variety of medical specialists represented in the survey (Table 1).
Table 1. Response rate, level of training, and medical specialties of respondents. Probabilities may not add up to 100% due to rounding.

<table>
<thead>
<tr>
<th>Level of training</th>
<th>Leiden University Medical Center (n=33), n (%)</th>
<th>Amsterdam University Medical Center (n=31), n (%)</th>
<th>Total (N=64), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intensivist</td>
<td>16 (48)</td>
<td>11 (36)</td>
<td>27 (42)</td>
</tr>
<tr>
<td>Intensive care unit fellow</td>
<td>5 (15)</td>
<td>6 (19)</td>
<td>11 (17)</td>
</tr>
<tr>
<td>Resident</td>
<td>12 (36)</td>
<td>14 (45)</td>
<td>26 (41)</td>
</tr>
<tr>
<td>Medical specialty</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internal medicine</td>
<td>11 (33)</td>
<td>7 (23)</td>
<td>18 (28)</td>
</tr>
<tr>
<td>Anesthesiology</td>
<td>11 (33)</td>
<td>16 (52)</td>
<td>27 (42)</td>
</tr>
<tr>
<td>Pediatric medicine</td>
<td>3 (9)</td>
<td>0 (0)</td>
<td>3 (5)</td>
</tr>
<tr>
<td>Emergency medicine</td>
<td>1 (3)</td>
<td>0 (0)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Pulmonology</td>
<td>1 (3)</td>
<td>2 (7)</td>
<td>3 (5)</td>
</tr>
<tr>
<td>Surgery</td>
<td>1 (3)</td>
<td>0 (0)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Neurosurgery</td>
<td>1 (3)</td>
<td>0 (0)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Neurology</td>
<td>1 (3)</td>
<td>0 (0)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Resident not in training</td>
<td>3 (9)</td>
<td>6 (16)</td>
<td>9 (14)</td>
</tr>
</tbody>
</table>

aThe response rate for this group was 33 of 40 (83%).
bThe response rate for this group was 31 of 53 (58%).
cIncludes physician assistants.

Current Decision-Making Behavior With Respect to Discharging ICU Patients

Responses on current discharge practices are visualized in Figure 1. Physicians disagreed on the complexity of the decision to discharge a patient from the ICU, with 23 of 64 (36%) agreeing or strongly agreeing with the Q1 statement and 22 of 64 (34%) disagreeing or strongly disagreeing (Table 2). A nonsignificant difference was observed between experienced intensivists and other physicians (Multimedia Appendix 3). For question 2, 61 of 64 (95%) of physicians agreed or strongly agreed that readmission was an important factor in the decision to discharge a patient. Besides a patient’s readmission risk, physicians indicated that bed availability was an important factor in their decision to discharge (47/64, 73%, Q3). Furthermore, we asked physicians to report their average certainty regarding their estimation of the readmission risk of a patient after discharge. The median certainty score was 7 (IQR 7-8) for the whole group, with no significant difference observed between the two locations (P=.79). Patient groups for which the decision to discharge was perceived to be most challenging included patients with a long length of ICU stay (44/64, 69%) and readmitted patients (44/64, 69%: Table 3). Multimedia Appendix 4 shows the open-answer questions regarding patient groups and clinically relevant patient factors in the decision to discharge. The most reported reason for a complex decision to discharge was case complexity (9/64, 14%). The most frequently mentioned factor influencing the decision to discharge a patient (in relation to the patient’s clinical state, process-related factors, and factors related to the receiving ward) was the level of care and facilities at the ward (23/64, 36%), followed by the general clinical state of the patient (17/64, 27%) and the patient’s alarm ability (10/64, 16%).
Figure 1. Responses to statements regarding current intensive care unit discharge practices. The bar width of the answers indicates the number of respondents that chose that option. Q1: “The decision to discharge a patient to a lower care ward is complex”; Q2: “A patient’s ICU readmission risk is an important factor in my decision to discharge”; Q3: “I take bed availability into account for my decision to discharge a patient.” LUMC: Leiden University Medical Center; UMC: University Medical Center. *P<.05.

Table 2. Responses to statements. *P values in italics represent a significant difference (P<.05) between the Leiden University Medical Center and Amsterdam University Medical Center respondents. Results are reported on a 5-point Likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree). Scores >3 indicate median agreement with the statement and results <3 median disagreement.

<table>
<thead>
<tr>
<th>Question</th>
<th>Total, median (IQR)</th>
<th>Leiden University Medical Center, median (IQR)</th>
<th>Amsterdam University Medical Center, median (IQR)</th>
<th>P valuesa</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Domain 1: Physicians’ current decision-making behavior with respect to discharging ICUb patients</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1: “The decision to discharge a patient to a lower care ward is complex”</td>
<td>3 (2-4)</td>
<td>3 (2-3)</td>
<td>4 (2-4)</td>
<td>.04</td>
</tr>
<tr>
<td>Q2: “A patient’s ICU readmission risk is an important factor in my decision to discharge”</td>
<td>4 (4-4)</td>
<td>4 (4-4)</td>
<td>4 (4-5)</td>
<td>.09</td>
</tr>
<tr>
<td>Q3: “I take bed availability into account for my decision to discharge a patient”</td>
<td>4 (3-4)</td>
<td>4 (3-4)</td>
<td>4 (3.5-4)</td>
<td>.43</td>
</tr>
<tr>
<td><strong>Domain 2: Physicians’ perspectives on the use of artificial intelligence (AIc)–based clinical decision support tools in general</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q4: “I am familiar with the concept of AI”</td>
<td>4 (4-4.25)</td>
<td>4 (4-4)</td>
<td>4 (4-5)</td>
<td>.004</td>
</tr>
<tr>
<td>Q5: “I believe AI could support me in my work as physician”</td>
<td>4 (4-4)</td>
<td>4 (4-4)</td>
<td>4 (4-4.5)</td>
<td>.006</td>
</tr>
<tr>
<td>Q6: “I believe that AI will take over my job in the future”</td>
<td>2 (2-3)</td>
<td>2 (2-3)</td>
<td>2 (2-2.5)</td>
<td>.22</td>
</tr>
<tr>
<td>Q7: “I believe AI understands my work sufficiently in order to support me”</td>
<td>3 (3-4)</td>
<td>3 (3-4)</td>
<td>3 (3-4)</td>
<td>.39</td>
</tr>
<tr>
<td>Q8: “I believe in the added value of AI based decision support at the ICU”</td>
<td>4 (4-4)</td>
<td>4 (4-4)</td>
<td>4 (4-4)</td>
<td>.41</td>
</tr>
<tr>
<td><strong>Domain 3: Physicians’ willingness to incorporate the discharge decision support tool in daily clinical practice</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q9: “An AI based decision support for ICU readmission could be of positive value in the decision to discharge a patient”</td>
<td>4 (4-4)</td>
<td>4 (4-4)</td>
<td>4 (4-4)</td>
<td>.02</td>
</tr>
<tr>
<td>Q10: “It is important for me to have insight in the contributing factors to the predicted chance of readmission”</td>
<td>4 (4-4.25)</td>
<td>4 (4-5)</td>
<td>4 (4-4)</td>
<td>.03</td>
</tr>
<tr>
<td>Q18: “I assume that no readmission risk prediction score could influence my behavior”</td>
<td>2 (2-2)</td>
<td>2 (2-2)</td>
<td>2 (2-2)</td>
<td>.11</td>
</tr>
<tr>
<td>Q19: “I’m willing to consult the prediction of the decision support tool before making my decision to discharge a patient”</td>
<td>4 (4-4)</td>
<td>4 (4-4)</td>
<td>4 (4-4)</td>
<td>.47</td>
</tr>
<tr>
<td>Q20: “Taking into account the current workload at my department, I have time to take in the prediction score provided by the decision support tool and to take this into account for my decision to discharge a patient”</td>
<td>4 (4-4)</td>
<td>4 (3-4)</td>
<td>4 (4-4)</td>
<td>.11</td>
</tr>
</tbody>
</table>

aP values were calculated with the Mann-Whitney U test.
bICU: intensive care unit.
cAI: artificial intelligence.
Table 3. Patient groups for which the decision to discharge was perceived as most challenging (one or more options could be chosen).

<table>
<thead>
<tr>
<th>Patient groups</th>
<th>Respondents, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long admission</td>
<td>44 (69)</td>
</tr>
<tr>
<td>Currently readmitted</td>
<td>44 (69)</td>
</tr>
<tr>
<td>Elderly</td>
<td>7 (11)</td>
</tr>
<tr>
<td>COVID-19</td>
<td>4 (6)</td>
</tr>
<tr>
<td>Other</td>
<td>12 (19)</td>
</tr>
</tbody>
</table>

Attitudes and Perspectives Toward AI-CDS Tools in the ICU

The respondents were familiar with the concept of AI (62/64, 97% agreed or strongly agreed with Q4) and the majority agreed that AI could support them in their work as a physician (55/64, 86% agreed with Q5; Figure 2). Respondents from the development site (Amsterdam UMC) were more familiar with the concept of AI ($P=.004$, Q4) and agreed more with the statement that AI could support them in their work as a physician ($P=.006$, Q5) than the LUMC respondents. The majority did not believe that AI would take over their job in the future (46/64, 72% disagreed or strongly disagreed with Q6), and the respondents were indecisive on whether AI understood their work sufficiently to support them (26/64, 41% agreed or strongly agreed and 12/64, 19% disagreed or strongly disagreed with Q7). Nevertheless, 55 of 64 (86%) respondents believed in the added value of AI-CDS in the ICU (Q8). The more experienced intensivists agreed significantly less with the statement “I believe AI could support me in my work as a physician” (Multimedia Appendix 4). This finding was compatible with the responses to Q7 and Q8, indicating that the more experienced respondents were less convinced that AI understood their work sufficiently and that AI could be of added value at the ICU.

Willingness to Incorporate a Discharge AI-CDS Tool in Daily Clinical Practice

The respondents agreed or strongly agreed that a discharge AI-CDS tool could be of positive value (59/64, 92%; Q9), and were willing to take the time to consult the AI-CDS and to take the prediction of the tool into consideration before discharging a patient (44/64, 69%; Q20 and 58/64, 91%; Q19; Figure 3). Furthermore, respondents disagreed or strongly disagreed with the statement “I assume that no readmission risk prediction score could influence my behavior” (53/64, 83%, Q18). Amsterdam UMC respondents agreed more with the statement “an AI based decision support for ICU readmission could be of positive value in the decision to discharge a patient” than the LUMC respondents ($P=.02$, Q9), but the difference was small. Q10 emphasizes the need of physicians for prediction tools to be explainable (57/64, 89% agreed or strongly agreed). This need was more important for the LUMC physicians ($P=.03$).
Figure 3. Statements regarding willingness to incorporate a discharge decision support tool in daily clinical practice. Q9: “An AI based decision support for ICU readmission could be of positive value in the decision to discharge a patient”; Q10: “It is important for me to have insight in the contributing factors to the predicted chance of readmission”; Q18: “I assume that no readmission risk prediction score could influence my behavior”; Q19: “I’m willing to consult the prediction of the decision support tool before making my decision to discharge a patient”; Q20: “Taking into account the current workload at my department, I have time to take in the prediction score provided by the decision support tool and to take this into account for my decision to discharge a patient.” LUMC: Leiden University Medical Center; UMC: University Medical Center. *P<.05.

Physicians were asked to indicate the threshold of predicted readmission and mortality risk (on a scale from 0 to 100) above which they would not discharge a patient to the ward, and the threshold below which they would discharge a patient (Figure 4). Results varied widely. The LUMC respondents reported that a median 40% (IQR 20%-50%) readmission and mortality risk or higher would cause them to postpone discharge, compared to a 20% (IQR 10%-30%) risk for the Amsterdam UMC group. The LUMC group indicated that a median readmission and mortality risk of 20% (IQR 10%-33%) or lower would be acceptable to discharge a patient, compared to a 10% (IQR 7.5%-20%) risk for the Amsterdam UMC group.

Figure 4. Predicted readmission and mortality risk that would influence physicians’ behavior in discharging or not discharging an intensive care unit patient. LUMC: Leiden University Medical Center; UMC: University Medical Center.

Desired Workflow for the Tool and End Users
A total of 40 of 64 (63%) of the ICU physicians preferred that risk prediction be integrated in EHRs, while 21 of 64 (33%) preferred a stand-alone dashboard. The moments that the respondents most often chose for the AI-CDS to be displayed were during morning handover (24/64, 38%), morning rounds (21/64, 33%), and grand rounds or bedside multidisciplinary consultations (28/65, 44%; Table 4). The respondents indicated that AI-CDS predictions, if they were not continuous, should be updated before these moments to be of value to the end users. The tool was indicated to be most relevant for supervisors (ie, responsible board-certified intensivists; 62/64, 97%), intensive care fellows (57/64, 89%) and residents (42/64, 66%; Table 5).
Table 4. Desired moment to display the prediction tool, with approximate times. More than one option could be chosen.

<table>
<thead>
<tr>
<th>Moments</th>
<th>Time</th>
<th>Respondents (N=64), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morning handover</td>
<td>7:45 AM</td>
<td>24 (38)</td>
</tr>
<tr>
<td>Before morning rounds</td>
<td>8:30 AM</td>
<td>6 (9)</td>
</tr>
<tr>
<td>Morning rounds</td>
<td>8:45 AM</td>
<td>21 (33)</td>
</tr>
<tr>
<td>Grand rounds or bedside multidisciplinary consultation</td>
<td>11:30 AM</td>
<td>28 (44)</td>
</tr>
<tr>
<td>Multidisciplinary consultation</td>
<td>2 PM</td>
<td>17 (27)</td>
</tr>
<tr>
<td>Evening rounds</td>
<td>4:15 PM</td>
<td>6 (9)</td>
</tr>
<tr>
<td>Daily care</td>
<td>All day</td>
<td>8 (13)</td>
</tr>
</tbody>
</table>

Table 5. End users for whom the tool was deemed to be most relevant. More than one option could be chosen.

<table>
<thead>
<tr>
<th>End users</th>
<th>Respondents (N=64), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bed coordinators</td>
<td>33 (52)</td>
</tr>
<tr>
<td>Supervisors, intensive care physicians</td>
<td>62 (97)</td>
</tr>
<tr>
<td>Intensive care fellows</td>
<td>57 (89)</td>
</tr>
<tr>
<td>Residents</td>
<td>44 (66)</td>
</tr>
<tr>
<td>Nurses</td>
<td>34 (53)</td>
</tr>
</tbody>
</table>

Open Comments

At the end of the survey, physicians could leave open comments and suggestions. Two physicians indicated a need for insights into what patient factors the predictions were based on to consider the tool safe and trustworthy. Besides the need for model explainability, a need for further validation of the tool before being able to trust it was mentioned. Furthermore, the combined outcome prediction (ie, readmission or mortality) was found to be problematic, with one physician expressing willingness to accept a high risk of readmission, but not mortality. Another comment was related to the finding that bed availability was important in the decision to discharge, as multiple physicians mentioned that they would accept a high risk of readmission if the decision freed a bed for a liver-transplant patient, for example.

Discussion

This study assessed the preimplementation of AI-CDS tools across 4 domains: physicians' current decision-making behavior regarding ICU discharge, their perspectives on AI, and their preferences for an AI-CDS tool's implementation and use in clinical practice. We found that nearly all ICU physicians were familiar with AI and had positive expectations, with 55 of 64 (86%) believing that AI could support them in their work as physicians. Not all physicians found the decision to discharge a patient complex, yet 59 of 64 (92%) agreed that a discharge decision support tool could be of value. Physicians at the site where the AI-CDS tool was developed showed greater familiarity with AI and had a stronger belief in the supportive role of AI in general, but also had a stronger belief that an AI-CDS tool specifically for discharge decision support would be useful compared to physicians at the nondevelopment site. Physicians from the nondevelopment site attached more importance to understanding which factors contributed to the predictions.

A positive attitude among physicians toward the use of the AI-CDS tool has also been found in other studies [15,16,18,20]. Interestingly, most respondents in our study believed in the added value of AI-CDS tools, while only 26 of 64 (41%) agreed or strongly agreed that AI understood their work sufficiently to support them. As in previous surveys [18,20], this incongruous finding could be explained by the fact that these physicians had not worked with AI-CDS tools when the study was conducted, and they therefore did not know if these tools were capable of capturing the complex ICU environment [28]. Lastly, the literature confirms the effect of bed capacity on physicians’ decision to discharge, which could limit the applicability of the AI-CDS tool in settings where bed capacity is low [21].

A recent scoping review of guidelines for the development of AI-CDS tools concluded that more focus on implementation strategy is needed for effective integration in the clinical setting [29]. Human-factors research, in the form of qualitative interviews and questionnaires, may enhance the uptake of AI-CDS tools, as this approach may improve the system’s design, training process, and implementation strategies [12,17]. We recommend focusing on the important local and sociotechnical context of each preimplementation site to meet the challenge of the “last mile” of implementation [11,21,22,30,31]. The positive attitudes and willingness to use AI-CDS tools we observed are positive indicators of the acceptance of this new technology [32,33], but they also underpin the idea that expectations should be aligned with the intended use of the AI-CDS tool to be adopted [17]. Moreover, it will be of value to repeat our questionnaire after the implementation of the AI-CDS tool for discharging ICU patients, as it has previously been observed that physicians showed
reduced excitement ($P<.01$) about AI-CDS after implementation [34].

As illustrated by the differences in familiarity and enthusiasm toward AI-CDS at the development and nondevelopment sites, sufficient attention should be paid to training and informing physicians on the use of the AI-CDS tool in their daily practice [10]. This training should also encompass the ethics and responsibilities of using AI-CDS in health care, as the physicians retain final responsibility for treatment decisions [33]. Lastly, training will be needed to educate physicians on the interpretation of the mortality or readmission risk predictions, as we observed a range of answers regarding the threshold at which patients would or would not be discharged to lower care wards (Figure 3). Due to a significant imbalance in the number of patients that were or were not readmitted or died after discharge, risk predictions are skewed along the 0% to 100% scale, being concentrated around an event rate of 5.3% [25]; the respondents were not informed of this. Therefore, attention should be paid to the interpretation of these calibrated risk predictions during training, as perceptions clearly differed on what constituted high and low risks for this outcome.

The implications of this study for the design process of AI-CDS tools include the need for explainable AI, as most respondents indicated a need to have insight into the factors contributing to “black box” predictions. We want to stress that addressing explainability is not the only factor required for a successful AI-CDS implementation; rather, the incorporation of domain expertise, the sociotechnical context, and physicians’ perspectives should be taken into account during the whole development, design, and implementation process [31,35,36]. We recommend that AI-CDS developers perform user and human-factor research in an early phase of design and development to maximize impact and smooth integration into the current decision-making process [11].

A limitation of the current study was that we only conducted the survey at 2 academic tertiary referral hospitals in the Netherlands. This could reduce the generalizability of our findings; for example, ICU physicians from nonacademic hospitals may be less familiar with AI. Secondly, the respondents may have had differences in their understanding of AI, as we did not provide a clear definition of AI to the end users in order to keep the questionnaire concise. Another limitation was that we did not formally assess the validity and reliability of this questionnaire. However, we did construct the questionnaire with a broad team of experts and performed a feasibility study at LUMC before generalizing the questions to be applicable to Amsterdam UMC. Future research could develop validated questionnaires for the preimplementation of AI-CDS tools, and the 4 domains presented here relating to current decision-making, workflow, and perspectives toward AI-CDS may serve as a blueprint. The increased workload caused by the pandemic may have impacted our response rate (64/93; 69%). However, the different levels of training and variety of medical specialties of physicians working at ICUs were represented in our sample of ICU clinicians, and few differences were observed between experienced ICU physicians and other respondents (Multimedia Appendix 3). Nevertheless, a nonresponse bias may have affected our results, as the clinicians that did fill in the questionnaire could have had a higher interest in AI-CDS compared to nonrespondents.

To conclude, this survey provides valuable insights into current decision-making behavior and perspectives on the use of AI-CDS tools that can be used in the implementation process and the training of end users. Positive attitudes were reported toward AI-CDS in general and for an AI-CDS tool for discharging ICU patients in particular, even though not all the respondents perceived the decision to discharge a patient to be complex. Observed differences between the 2 study sites, which had different levels of involvement in the development of the AI-CDS tool, show the need for education and training in departments with little experience with AI-CDS. We recommend that developers of AI-CDS tools involve their end users early in the design process and perform preimplementation by means of surveys to investigate potential acceptance in the local context, improve the system’s design and clinical workflow design, and ultimately facilitate clinical uptake.

Acknowledgments

The authors would like to thank Bas Becker, Maurits Dekker, and Aletta de Beer for their contribution to the development of the questionnaire. For this project, the Leiden University Medical Center and Pacmed authors received funding for their public-private partnership from Top Sector LSH.

Authors’ Contributions

SVDM, AAHDH, GC, PJT, and MSA contributed to concept and design. SVDM, PJT, and MSA contributed to acquisition of data. SVDM, AAHDH, MSA, GC, IMJK, and EWS contributed to analysis and interpretation of data. SVDM and MSA drafted the manuscript. SVDM, EWS, AAHDH, IMJK, GC, PJT, and MSA contributed to critical revision of the paper for important intellectual content and approved the final manuscript. SVDM and MSA contributed to the statistical analysis. MSA, EWS, GC, and IMJK supervised the project.

Conflicts of Interest

GC was an employee of Pacmed during this study. PJT received royalties from Pacmed for the Amsterdam University Medical Center during this study. SVDM is an employee of Healthplus.ai and discloses having received funding from The European Partnership from Top Sector LSH.
Regional Development Fund. The publication of the results was not conditional on approval from Pacmed, Leiden University Medical Center, or Amsterdam University Medical Center. No other disclosures are reported.

Multimedia Appendix 1
Rationale per question in the questionnaire.

Multimedia Appendix 2
Original questionnaire.

Multimedia Appendix 3
Statement questions: intensivists versus other participants.

Multimedia Appendix 4
Open answer questions.

References


31. van der Meijden et al. JMIR HUMAN FACTORS 2023 | vol. 10 | e39114 | p.780 https://humanfactors.jmir.org/2023/1/e39114


Intensive Care Unit Physicians' Perspectives on Artificial Intelligence–Based Clinical Decision Support Tools: Preimplementation Survey Study

Please cite as:
van der Meijden SL, de Hond AAH, Thoral PJ, Steyerberg EW, Kant IMJ, Cinà G, Arbous MS
Intensive Care Unit Physicians' Perspectives on Artificial Intelligence–Based Clinical Decision Support Tools: Preimplementation Survey Study
JMIR Hum Factors 2023;10:e39114
URL: https://humanfactors.jmir.org/2023/1/e39114
doi: 10.2196/39114
PMID: 36602843

©Siri L van der Meijden, Anne A H de Hond, Patrick J Thoral, Ewout W Steyerberg, Ilse M J Kant, Giovanni Cinà, M Sesmu Arbous. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 05.01.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
The Perceived Ease of Use and Perceived Usefulness of a Web-Based Interprofessional Communication and Collaboration Platform in the Hospital Setting: Interview Study With Health Care Providers

Jason Xin Nie1, MSc; Christine Heidebrecht1, MSc; Andrea Zettler1, BA; Jacklyn Pearce2, MSc; Rafael Cunha1, B'Tech; Sherman Quan3, MSc; Elizabeth Mansfield1, PhD; Terence Tang1,4, MSc, MD

1Institute for Better Health, Trillium Health Partners, Mississauga, ON, Canada
2Bellwoods Centres for Community Living Inc, Toronto, ON, Canada
3Accenture, Toronto, ON, Canada
4Department of Medicine, University of Toronto, Toronto, ON, Canada

Abstract

Background: Hospitalized patients with complex care needs require an interprofessional team of health professionals working together to support their care in hospitals and during discharge planning. However, interprofessional communication and collaboration in inpatient settings are often fragmented and inefficient, leading to poor patient outcomes and provider frustration. Health information technology can potentially help improve team communication and collaboration; however, to date, evidence of its effectiveness is lacking. There are also concerns that current implementations might further fragment communication and increase the clinician burden without proven benefits.

Objective: In this study, we aimed to generate transferrable lessons for future designers of health information technology tools that facilitate team communication and collaboration.

Methods: A secondary analysis of the qualitative component of the mixed methods evaluation was performed. The electronic communication and collaboration platform was implemented in 2 general internal medicine wards in a large community teaching hospital in Mississauga, Ontario, Canada. Fifteen inpatient clinicians in those wards, including nurses, physicians, and allied health care providers, were recruited to participate in semistructured interviews about their experience with a co-designed electronic communication and collaboration tool. Data were analyzed using the Technology Acceptance Model, and themes related to the constructs of perceived ease of use (PEOU) and perceived usefulness (PU) were identified.

Results: A secondary analysis guided by the Technology Acceptance Model highlighted important points. Intuitive design precluded training as a barrier to use, but lack of training may hinder participants’ PEOU if features designed for efficiency are not discovered by users. Organized information was found to be useful for creating a comprehensive clinical picture of each patient and facilitating improved handovers. However, information needs to be both comprehensive and succinct, and information overload may negatively impact PEOU. The mixed paper and electronic practice environment also negatively impacted PEOU owing to unavoidable double documentation and the need for printing. Participants perceived the tool to be useful as it improved efficiency in information retrieval and documentation, improved the handover process, afforded another mode of communication when face-to-face communication was impractical, and improved shared awareness. The PU of this tool depends on its optimal use by all team members.

Conclusions: Electronic tools can support communication and collaboration among interprofessional teams caring for patients with complex needs. There are transferable lessons learned that can improve the PU and PEOU of future systems.
health information technology; communication and collaboration; teamwork; design; technology acceptance model; qualitative method; communication; collaboration; hospital; care; team; professional; support; health information; technology; clinician; members; complex; lesson; education

Introduction

Background

Patients with complex care needs admitted to hospitals often require the services of an interprofessional team of health professionals working together to support their care [1]. However, in inpatient settings, interprofessional communication is often fragmented and inefficient [2,3]. Poor communication and teamwork can contribute to poor patient outcomes, such as delayed discharge, medication errors, and adverse and sentinel events, including death [4-7]. It can also lead to frustration among health care providers [8], especially when the providers are not on the same page regarding the plan of care [9,10].

Health information technology has the potential to improve interprofessional communication in hospital settings. Communication technology tools that are used vary between and within institutions and can range from numeric pagers to mobile devices or specialized software applications with varying degrees of integration with electronic health records [11-15]. Common concerns with existing technology include lack of context and structure, interruptive nature, privacy and security concerns, and a lack of visibility to the entire care team [12,15-20]. The information required to best address a patient with complex care needs may also exist in a combination of paper and disparate electronic systems, resulting in various team members being unaware of or unable to access information critical to providing the best quality of care in a timely and efficient manner. Systematic reviews published in 2012 and 2019 highlight the lack of high-quality evidence on the effectiveness of current communication tools in hospital settings [12,14]. Moreover, there are concerns that these technologies are not optimally designed, and their use may further fragment communication and increase the demand on clinicians without demonstrating benefits [21].

To generate transferrable lessons that may improve the design of future health information technology solutions aimed at facilitating communication and collaboration between clinicians of interprofessional teams within hospitals, we performed a secondary analysis [22,23] of qualitative data collected as part of a mixed methods evaluation of a co-designed interprofessional communication and collaboration tool [24]. Results from the mixed methods study showed improved teamwork (encompassing both communication and relational aspects) in one of the two study wards after the introduction of the tool, without meaningful changes in face-to-face communication patterns during team rounds or adverse events in both wards. There is potential for an electronic tool to improve teamwork and communication, but success is dependent on the complex interactions of technological and nontechnological factors [24]. The focus of this paper is to analyze our qualitative data using Technology Acceptance Model (TAM) to understand clinicians’ perspectives on the tool’s perceived usefulness (PU) and perceived ease of use (PEOU) to generate lessons relevant for the design of future interventions.

An Overview of the Electronic Communication and Collaboration Platform

To improve communication and collaboration among the interprofessional teams at the hospital, our team used agile methodology and co-designed a web-based technology platform with frontline clinicians using a variety of design methods as described by Tang et al [24,25]. It addresses issues with handoffs (with a physician sign-out tool), interprofessional collaboration (through the interprofessional care planner where information relevant to the team from each discipline can be viewed in one place and the patient flow planner in which barriers to discharge are identified and tracked), and team communication (secured team messaging that is attached to a patient and viewable by the entire care team). It also evolved to include an electronic discharge summary and an associated patient-oriented discharge summary to facilitate care transitions. Although the focus was on communication and collaboration, a progress note module (where typed notes can be generated and printed for the paper chart) was also developed to facilitate workflow and reduce double documentation.

The tool is a web-based platform that, although distinct from the hospital’s primary vendor health information system (HIS), can retrieve information from and write information to the primary HIS using Health Level Seven, a technical standard that allows health-related information to be exchanged between health care applications [26]. It does not replace but augments the HIS by providing communication and collaboration features designed to fit the clinician workflow. The architecture of the Care Connector is modular, allowing each module to be developed independently while addressing different yet interconnected clinical workflows. The 6 key modules (Figure 1) are interconnected with each other and with the primary HIS by sharing information, thereby allowing for continuity of information (as changes in one module are reflected in other modules and in the HIS in real time), reducing the need for repeated data entries or the likelihood of missed information. It also allows information reuse (eg, past medical history captured in the physician sign-out is reused in the interprofessional care planner and discharge summary) to improve communication and reduce documentation effort.

https://humanfactors.jmir.org/2023/1/e39051

 JMIR Hum Factors 2023 | vol. 10 | e39051 | p.783

KEYWORDS

health information technology; communication and collaboration; teamwork; design; technology acceptance model; qualitative method; communication; collaboration; hospital; care; team; professional; support; health information; technology; clinician; members; complex; lesson; education
The TAM, first developed in 1985 by Fred Davis, is used to provide a theoretical basis “of the effect of system characteristics on user acceptance of computer-based information systems” [27,28]. The TAM theorizes that actual system use is determined by a potential user’s attitude toward using the system, which in turn is based on the following 2 key beliefs: PU and PEOU [29]. PU is defined as “the degree to which an individual believes that using a particular system would enhance their job performance.” PEOU is defined as “the degree to which an individual believes that using a particular system would be free of physical and mental effort.” Moreover, PEOU is hypothesized to have a causal effect on PU because the easier a system is to use, the more useful the user will find the system [29].

**Methods**

**Study Design**

A secondary analysis was performed on the qualitative component of a mixed methods study conducted between February 2016 and July 2017 to assess the impact of an electronic communication and collaboration tool on communication, teamwork, and adverse events [24].
Ethics Approval
Ethical approval was obtained from the Research Ethics Board of Trillium Health Partners (approval number: ID#691).

Participants and Setting
Trillium Health Partners is one of the largest community-based hospital systems in Canada with 1306 beds across 3 sites. An electronic communication and collaboration tool, described in the section above, was implemented in 2 of the 5 General Internal Medicine wards at the Credit Valley Hospital site. Nurses and allied health care staff were ward based, whereas physicians who provided care to patients were dispersed throughout the hospital. At the time of the study, patient information was split between the hospital HIS and paper charts where progress notes and documentation were noted. In our mixed methods study, we recruited a diverse sample of frontline health care personnel using a purposeful maximum variation sampling strategy [30]. Potential participants in clinical and logistical roles in the 2 General Internal Medicine wards where the electronic tool had been deployed and used for at least 6 months were invited to participate.

Data Collection and Analysis
CH recruited, acquired consent, and interviewed all the participants. A copy of the full interview guide is presented in Textbox 1. EM, TT, CH, AZ, and JXN were engaged in the analysis. An inductive approach was used in this study. Three researchers (AZ, TT, and CH) independently reviewed a purposive sample of 4 transcripts and, during a series of meetings, developed a coding framework. Subsequently, 2 (CH and AZ) researchers coded all the transcripts, with each member being the primary coder for half the transcripts, and second coded the other half to ensure that the codes were applied appropriately and consistently. The team resolved issues, came to consensus via discussions and meetings, and then reviewed the coded data and identified key themes. CH sent all the participants an email of a summary of major findings for member-checking to which no participant objected. All authors contributed to the writing of the manuscript.
Textbox 1. Interview guide.

Background information

- Tell me a bit about your work and clinical role at the hospital.
- What other experiences have you had working with communication systems similar to Care Connector?
- How comfortable are you with information technology in general?
- How long have you used Care Connector?
- How often do you interact with Care Connector? (i.e. daily; per shift; weekly)
  - When in the day, or during your shift, do you tend to interact with Care Connector?
- When do you tend to interact with other care providers to make plans for patient care? Do these interactions involve Care Connector?

Impact of Care Connector on workflow, patient care, and interprofessional relations

- What modules do you primarily use?
- What gaps do you see Care Connector as addressing? (quality of patient care; interprofessional communications, patient handover, workflow efficiency etc.)
- Has Care Connector affected your workflow? If so, how? Provide an example/story illustrating this.
- Has using Care Connector affected patient care? If so, how? Provide an example/story illustrating this.
- Has Care Connector affected your workflow? If so, how? Provide an example/story illustrating this. Has using Care Connector affected patient care? If so, how? Provide an example/story illustrating this.
- Has using Care Connector affected your communications with other (physicians/nurses/allied health professionals/unit clerks/flow team members: insert appropriate role depending on interviewee’s role)? If so, how? Provide an example/story illustrating this.
- Has using Care Connector affected your communications with other team members? [Specify physicians, nurses, allied health professionals unit clerks/flow team members, excluding the interviewee’s role, which has been covered above] If so, how?
- Has Care Connector affected the relationship between staff/health care professionals? If so, how?
- Has Care Connector affected teamwork between you and your colleagues? If so, how?
- How do you feel about teamwork between you and other (choose discipline depending on role of respondent: physicians/allied health/nurses/etc.)? What about with other hospital staff?
- What are your thoughts on the effectiveness of care rounds?
  - “Does Care Connector support you in any way at care rounds?
  - Describe and get their feedback on the idea of the marketplace

Strengths and challenges of working with the new Care Connector modules

- How did you find the process used to introduce, implement and obtain feedback about Care Connector?
  - What worked well?
  - What could be improved?
- What features of the Care Connector modules do you find most useful?
- What features of the Care Connector modules need improvement?
- What challenges from a workflow and clinical documentation perspective has using Care Connector created, if any?
- If there were times when you had a choice between using Care Connector and completing a task using a conventional approach (e.g. when documenting progress notes), what made you choose Care Connector over the traditional approach or vice versa?
- Are there any unintended benefits or consequences you discovered from using Care Connector?
- Why or why not should Care Connector be introduced to other departments and hospital units?
- What are some other healthcare settings where Care Connector might be useful?
- If there were to be a module, or multiple modules, that would involve patients – and would facilitate communication between team members and patients themselves – do you think that that would be valuable?

Conclusion

- Is there anything else that you would like to comment about that I haven’t asked you about?
Secondary Analysis

Our team performed a secondary analysis of all the original transcripts using the TAM lens by mapping questions from the original interview guide that were relevant to the TAM model (Textbox 1). Interview questions related to PEOU included the following: “What features of the technology need improvement?” “If there were times when you had a choice between using the platform and completing a task using a conventional approach, what made you choose the platform over the approach or vice versa?” “What challenges from a workflow and clinical documentation perspective has using the platform created, if any?” The participants were asked which features or functionalities were easy or difficult to use. Responses were analyzed to identify comments related to PEOU (eg, confusion, frustration, ease, difficulty, and intuitiveness). Regarding PU, we asked the following questions: “What modules do you primarily use?” “What features of the technology do you find most useful?” “What gaps do you see the platform addressing?” “Has the technology affected your workflow?” Finally, we specifically explored the perceived role of technology in facilitating teamwork and communication in team-based care as part of understanding PU. The questions included the following: “Has using the technology affected your communications with other team members?” “Has the technology affected the relationship between healthcare providers?” “Has the platform affected teamwork between you and your colleagues?” “Has using the platform affected patient care?”

A thematic content analysis approach was applied [31]. Our coding methods have been described by Tang et al [24]. Key themes and relationships between the themes were identified inductively through team members’ individual reviews and group dialogue regarding code reports and memos. Themes related to the PEOU and PU of the TAM were used for the analysis.

Results

Overview

In total, 15 transcripts were included in this secondary analysis, including the perspectives of physicians (4/15, 27%), nurses (5/15, 33%), allied health care professionals (4/15, 27%), and nonclinical support personnel (2/15, 13%). Here, we report the findings of our secondary analysis from the perspective of the TAM (Figure 2). Using this focused analytic approach, the following themes emerged in relation to PEOU: learnability, information organization, functionality gaps discovered after deployment, and challenges related to the coexistence of paper and electronic systems. The following themes emerged in relation to PU: efficiency in information retrieval, improved handover processes, improved communication and teamwork, and the potential for improved shared awareness.

Perceived Ease of Use

With respect to PEOU, the following themes emerged: learnability, information organization (the need to achieve comprehensiveness without causing information overload), and your colleagues?” “Has using the platform affected patient care?”

A thematic content analysis approach was applied [31]. Our coding methods have been described by Tang et al [24]. Key themes and relationships between the themes were identified inductively through team members’ individual reviews and group dialogue regarding code reports and memos. Themes related to the PEOU and PU of the TAM were used for the analysis.

Results

Overview

In total, 15 transcripts were included in this secondary analysis, including the perspectives of physicians (4/15, 27%), nurses (5/15, 33%), allied health care professionals (4/15, 27%), and nonclinical support personnel (2/15, 13%). Here, we report the findings of our secondary analysis from the perspective of the TAM (Figure 2). Using this focused analytic approach, the following themes emerged in relation to PEOU: learnability, information organization, functionality gaps discovered after deployment, and challenges related to the coexistence of paper and electronic systems. The following themes emerged in relation to PU: efficiency in information retrieval, improved handover processes, improved communication and teamwork, and the potential for improved shared awareness.

Perceived Ease of Use

With respect to PEOU, the following themes emerged: learnability, information organization (the need to achieve comprehensiveness without causing information overload), and your colleagues?” “Has using the platform affected patient care?”

A thematic content analysis approach was applied [31]. Our coding methods have been described by Tang et al [24]. Key themes and relationships between the themes were identified inductively through team members’ individual reviews and group dialogue regarding code reports and memos. Themes related to the PEOU and PU of the TAM were used for the analysis.

Results

Overview

In total, 15 transcripts were included in this secondary analysis, including the perspectives of physicians (4/15, 27%), nurses (5/15, 33%), allied health care professionals (4/15, 27%), and nonclinical support personnel (2/15, 13%). Here, we report the findings of our secondary analysis from the perspective of the TAM (Figure 2). Using this focused analytic approach, the following themes emerged in relation to PEOU: learnability, information organization, functionality gaps discovered after deployment, and challenges related to the coexistence of paper and electronic systems. The following themes emerged in relation to PU: efficiency in information retrieval, improved handover processes, improved communication and teamwork, and the potential for improved shared awareness.

Perceived Ease of Use

With respect to PEOU, the following themes emerged: learnability, information organization (the need to achieve comprehensiveness without causing information overload),
Learnability: Trade-off Between Intuitive Design and Need for Learning

Through the agile software development methodology and continuous user engagement involved in its development [25], the platform was designed to be inherently intuitive, requiring minimal user training. Our findings indicated that these objectives were achieved; overall, the participants found the electronic communication and collaboration tool easy to use and adopt:

Yeah, I think it was pretty easy to pick up. You know it’s not hard to learn. And I think just a colleague of mine just showed me and it was fine. [Physician 1]

The implementation of the electronic tool, for me, I felt like it wasn’t too bad on learning stuff. It was pretty straightforward in terms of accessing, updating the information on there. [Nurse 4]

However, the ease of use could be hindered if the users had not discovered a particular system functionality. For example, a patient’s medical condition in the past medical history section were captured discretely on the platform (ie, requiring each condition to be entered into a separate row) to facilitate information reuse throughout the platform. Users had the option of clicking the “Add” button to add a new row to the list, but the electronic tool also allowed the user to simply hit the Enter key to get to the new line so that users were not slowed down by using the mouse when they had to enter a long list of medical conditions. However, users who were unaware of this feature found the system to be labor intensive:

It takes forever to type all these things out and you have to do it for each patient. And, I think each time you have to click something on the screen to make something happen, it just increases the amount of work you have to do. So, for example, the past medical history section has this, like, Add thing where you do one past medical history, add each one at a time. I would never use that...Whatever it is, I’m not going to keep going back and pressing “Add”. It seems to be a cognitive load I don’t need to deal with. [Physician 4]

Information Organization: Tension Between Comprehensiveness and Information Overload

The participants had different perceptions regarding the PEOU of the layout and information organization of the platform. The participants felt that the information in the electronic tool was more organized and consistent. Moreover, participants found that information is well organized and easy to locate is useful for creating a comprehensive clinical picture of each patient and facilitating improved handovers, demonstrating that improved PEOU is associated with increased PU:

Things are less easily missed perhaps...[In the electronic tool] you always have a consistent layout, and people tend to put information in the same area. So you know hopefully that you’re not going to miss a piece of information elsewhere, because it’s more consistently used amongst nursing staff. [Nurse 3]

I just feel that people are, it just seems to be more clear in the documentation being done in the [the electronic tool], I can’t explain why that is but it just is...The plan is better organized or the next steps are better listed. [Allied health 4]

However, electronic systems can contribute to information overload, and important information and day-to-day changes can sometimes be overlooked. Some participants suggested that using headings better, reducing the amount of scrolling, and highlighting key information or changes may improve PEOU. For example, a physician observed that a key component of the daily progress note, the physician’s impression and plan for the day, is sometimes difficult to locate because of content that was copied forward:

When I see progress notes that have come out of [the electronic tool], the problem I always have is that they all look the same and they don’t highlight the day’s problems as well...it’s sort of hidden in the body. You have to sort out what’s changed and the problem list still stays the same very often either because people don’t want to change it or again, there just wasn’t enough to change. It always involves sort of hunting and trying to see what is different in today’s note versus the note that was written yesterday, and trying to find the data that looks different to find out what happened. [Physician 2]

Similarly, another participant found that the tool could be improved by more prominent visual cues to highlight important information, especially for patients with complex medical needs:

I feel like the whole layout when you first open it and you have to like scroll through looking at all the different aspects, like where they’re from, if they’re diabetic, like how they take their meds [...], I feel like it’s so like...how do I describe this...like one colour, like nothing really stands out, I feel like it looks so...like, not blah, but it’s hard to find things if you’re trying to scroll through there fast. I feel like it could be more like friendly, like maybe more colours or like the way it’s laid out. [Nurse 2]

Functionality Gaps Discovered After Deployment: Potential to Improve PEOU

Several participants identified functionality gaps and workflow requirements that were previously unaccounted for, which limited the PEOU of the electronic platform. For example, the messaging component was a core feature of the system that allowed users to send messages to any member of the care team by name. However, clinicians might not always know the name of the team members they were sending the message to, but they did know the provider role they were trying to connect with. One participant suggested that having the added functionality to send a message to select roles within the patient’s care team would improve PEOU:

Well, I mean, I’m not completely clear on who I can send a message to. But I’m assuming you can send it to the allied health team but, you know, there is no part on [the electronic tool] which identifies who the
Coexistence of Paper and Electronic Systems Resulted in Workflow Challenges

The clinical practice environment of our organization at the time when the study was conducted was a mix of paper and electronic systems that clinicians had to navigate. Clinical documentation (eg, notes) was paper based, while some information (eg, vital signs and diagnostic testing results) was captured electronically. To reduce double documentation (ie, having users enter the same information both on paper and electronically on the communication tool), a co-designed feature of our system allowed users to efficiently generate documentation electronically. However, owing to the practice environment, this electronically generated documentation still needed to be printed and placed in paper charts. This administrative burden caused frustration for users and significantly limited the PEOU. Some users weighed the cost and benefit of the extra effort required for printing and reported that they would only use the tool when the benefits outweigh the time and effort required:

I think if I didn’t have to print the notes out and then put it in the chart, that definitely [would make me more likely] to chart things on the computer. Yeah, I think the main thing is I have to get the chart anyways, so sometimes it’s much faster for me to just scribble notes in the chart, whereas with [the electronic tool] I have that extra step of finding a computer, print it, find a printer, print, and then find the chart and putting it in the chart. [Physician 1]

In a mixed paper and electronic environment, participants often chose the method that was efficient for them in the moment for a particular task:

So one example is, for example, so if I have a longer note to type, like a family meeting that I need to document, I would probably use [the electronic tool], just because it’s a longer note and it would require more handwriting, if I were to write it out. So I would choose [the electronic tool] to document longer progress notes. In terms of handwriting, if I were to physically write in the chart it would be something very short. [Allied health 2]

In addition, although the system was designed to reduce double documentation, the paper documentation requirement had made this unavoidable in some situations. For example, an allied health professional expressed frustration with a specific assessment form that was not supported by the electronic tool:

I think one of the things specifically to me that I find a little bit frustrating in my work is kind of the double charting that we do. So basically we have an [assessment form] that we fill out for every new assessment that has all the information on the patient’s background and then what we found in the assessment and what our recommendations are. And then in addition to that we also do a chart note. So when I get a new assessment I have to do a new chart note, I have to do [an assessment form], I have to do an order in the chart. And I have to put a sign above their bed with my recommendation. So it’s a lot of double charting or double writing. [Allied health 1]

Perceived Usefulness

Regarding the PU and the role of technology in supporting teamwork and communication, the following themes emerged: efficiency, improved handover, mode of communication (electronic tools play a role when face-to-face communication is not possible), degree of use (usefulness depending on extent of use), and shared awareness (even in the absence of direct communication).

Efficiency in Information Retrieval and Documentation

The care planning module (distinct from documentation) of the electronic tool made it easier to retrieve information for care and planning. This section of the tool was primarily used by nursing and allied health care staff members. Key information necessary for care planning and decision-making was well organized under clear section headings without the need for reading through voluminous documentation, thus saving time and increasing efficiency. Allied health care personnel, who often found following physician notes challenging, appreciated that information was organized around headings that were relevant to care planning, which helped to make the information easy to find and actionable:

And I think that [electronic tool] gives a standardized format of how to, I mean (certain allied health disciplines) tend to have a standardized way of documenting whereas I find the physicians not always. So I find that...to read what their plan is...is easier. [Allied health 4]

It shortens the time that I have to spend digging for information because I have information readily accessible and available in some degree. [Nonclinical support 1]

It’s just better...the information is definitely more organized. And just the key things that we’re looking for it’s just...they’re all included in [the electronic tool], so it’s just easy for us to communicate. [Nurse 1]

Efficiency also increased through features supporting general documentation, including the ability to import information across different modules of the system and previous notes, as noted by a user:
I think that I have enjoyed the efficiency that it’s given me, and particularly with that import last note function, and I think a lot of us have used that because essentially we’re assessing the same sorts of things with patients every time we see them. We’re just updating, you know, their new functional status. And so it takes a lot of time to rewrite all of that, or if there’s specific things about the patient’s background or history that you want to mention, you don’t have to retype or rewrite all those things. So I think from an efficiency perspective, especially with the import last note function, it’s given me a lot more efficiency. For me, I can see a lot more patients because I’m not handwriting notes all day. So I can type faster than I can handwrite. So just from that perspective that’s been nice. [Allied health 1]

Improving Handover Processes

Participants uniformly perceived the electronic tool to be especially useful during handover (which occurred when physicians rotated off clinical service or provided weekend coverage and when nurses changed shifts). They perceived the communication tool as providing a structure for the handover process and reducing the likelihood of missed information:

> It provides a better hand over than we were doing before. I would always worry—I mean I tended to be pretty thorough in my emails and that but you would always worry that there were details that were missed, and email is just free form so it’s nice to have the organization the way it is now in terms of their past history, their issues and then the problem list, and so I think that’s probably a safer way to ensure that relevant information gets passed on. [Physician 2]

Another participant highlighted increased awareness of patient history because information in the tool was contributed by all previous providers rather than just from the previous shift, making the collective knowledge of the patient available:

> There’s a way for information to be passed on not just between nurses that are handing over but from prior nurses as well, because we can provide historical information on there to guide care. So I think there is more continuity in terms of information being passed forward, not just based on one shift’s information, but the information coming from many nurses prior to that. [Nurse 3]

Improved Communication and Teamwork With Team Members Not Physically Present on Unit

At our institution, physicians attended to patients in many different wards, whereas nursing and allied health care teams were assigned to one ward. Participants observed that face-to-face communication when engaging in active care planning is preferred whenever possible. Therefore, the electronic tool was particularly useful for communication and facilitating teamwork when team members are unable to see each other face-to-face. Allied health care participants commented on the improved quality of communication with physicians as they may not always be on the ward:

> I feel like [the electronic tool] might have made [teammwork] easier with the physicians. [...] We can also just look at what the physicians have written about the patient and their plans which can also limit the amount of time nursing is paging the physicians going, you know, “What do you want to do with this?” when they’ve probably already written it somewhere. [Nurse 2]

> [The electronic tool] sort of started to address the communication issue that we all sort of seem to have, communicating with the physicians I guess is what I’m referring to most. Again, because they’re not always on the unit, whereas the other staff, if we need to communicate with them, we can usually find them pretty easily. [Allied health 1]

Potential for Improved Shared Awareness Among Interprofessional Team Members

Participants reported that distinct from the ability to facilitate direct communication (eg, via messaging), the designed system was useful for teamwork and collaboration because it improved shared awareness among the team. Participants noted that as all team members had access to the tool, it was easy for all team members to be “on the same page” and understand shared goals for the patient:

> In a way, yes [the electronic tool addresses gaps in respondent’s work], because at least everyone that’s involved with the patient has access to it, so instead of it just being me trying to make those adjustments—you know, like allied health they have access and they can make those changes as well—so that definitely helps to put all the connections into place and stuff. [Nurse 4]

Moreover, the Care Planner module allowed all team members to contribute to the patients’ care plan. Understanding and contributing to the shared goals for the patient was identified as having the potential for more efficient discharge planning:

> It’s definitely improved communication. I know that some of the social workers, patient flow and, [nursing], we communicate through the [Care Planner]. So it just helps improve communication and then the discharges happen faster. There’s not a lot of gaps that we’ve missed with regards to discharge planning. So it definitely fills those communication gaps. [Nurse 1]

Although participants have outlined the many positive benefits of an electronic communication and collaboration system, PU is dependent on optimal use by all clinicians. When clinicians consistently use it, information is up-to-date and relevant. However, if this is not the case, the PU of the system will decrease as echoed by a participant:

> [My] only wish is [that] all the physicians were updating it. Because you know, some of them are better than others. Some of them are, you know, like updating daily or every other day or, you know, putting some extra notes and taking summaries...
Discussion

Principal Findings

This study explored clinicians’ perspectives on the PEOU and PU of a web-based electronic communication and collaboration platform designed to facilitate team-based care for hospitalized patients with complex needs. Our results demonstrate a number of transferable lessons for others designing and implementing health information technology aimed at facilitating team communication and collaboration for inpatient care.

The design goal of the platform was to be intuitive to users requiring minimal documentation or training. However, there are often trade-offs in the design of an intuitive user interface [32]. The situation experienced by our user, who was frustrated with having to click the “Add” button to add a text field while failing to recognize that pressing Enter key would do the same thing, highlighted trade-offs between affordance (the intuitiveness of a visual element), learnability, efficiency, and discoverability. The Add button had high affordance, leading to higher platform learnability, but it decreased efficiency (ie, the need for mouse click). Pressing Enter key was a more efficient way of accomplishing the task, but it had low affordance (no visual element) and required users to discover this feature [32,33].

More information is not always better especially as it pertains to health information technology. Information overload in electronic health records has been shown to contribute to clinician stress and burnout, worsened workloads, and create opportunities for errors [34,35]. Although the users of our electronic platform reported that succinct information that was well organized and comprehensive made the system easy to use, the sheer amount of information captured on the platform could detract from the ease of use of the system [35]. This might be due to repetitive information in progress notes made possible by import features, causing “note bloat” (unnecessarily long cut-and-pasted progress notes) [35] and remains a tension to be considered in future system designs.

Functionality and workflow gaps that were not identified in the co-design of the platform emerged after clinical use. In our example, it was the inability to send messages to professional roles when the name of an individual team member was not known, and the inability to search patients by name. Therefore, it is critical to periodically evaluate functionality and workflow after implementation to identify areas of improvement that were not initially foreseen.

Clinical environments are complex, and the combination of paper and disparate electronic systems presents a unique challenge to system designers. Our study highlighted that in these blended environments, printing and double documentation are major issues that designers should seek to eliminate.

Our findings indicate that an electronic communication and collaboration system can achieve high usefulness with respect to improving efficiency and supporting improved handover when it is appropriately adopted. Participants in our study reported an improved quality of patient handover with the electronic platform, as information transfer was standardized and important details were not missed, ensuring continuity of patient care. This is consistent with previous literature in which electronic sign-out or handoff tools have been shown to improve the process, with fewer information omissions and improved efficiency [36-38].

Our results showed that clinicians see value in face-to-face interactions in care planning, and electronic tools may play a role in situations where these interactions are not possible. A previous study observed that non-IT communication was positively correlated with software adoption, suggesting that the relationship was not substitutive but rather complementary [39]. Our findings also suggest that the use of electronic platforms complements, rather than substitutes, in-person communication; it has strengthened electronic information exchange, especially across disciplines and during handover; however, face-to-face interactions remain highly valuable in active care planning.

Although the TAM suggests that PEOU and PU predict actual system use, our results also suggest that actual system use may impact PU. It is not difficult to see that for a teamwork and collaboration tool, the lack of users can be a big reason why it can fail to live up to its intended purpose. Our data reinforce that a shared communication and collaboration tool is most effective when all team members use it and keep information up-to-date, which leads to increased system use by other members of the team, and increased system use may further increase information quality. Conversely, the lack of system use by other team members adversely affects the PU for those who use it. The design of team collaboration tools should, therefore, also include looking at the groups (rather than individuals) who will be using the tool (eg, using collaboration personas) [40,41]. This also has implications for the implementation strategy, as a limited deployment of such tool may limit its usefulness.

Finally, our data suggests that electronic tools can play a role in supporting shared awareness among teams. Common ground—shared knowledge and understanding among ≥2 agents—plays an essential role in communication and collaboration within health care teams [42-44]. Discharge, care planning, workflow planning, rounds, and patient goal setting have been observed as care areas in which the establishment of common ground is critical [44]. In addition to trust in colleagues’ abilities—and knowledge of limitations in ability—Kuziemsky et al [44] identify the “push” and “pull” of information exchange as key elements in the establishment of common ground. Technology can be designed to support grounding through electronic channels when face-to-face interactions are not feasible [44]: we observed that many of the platform’s modules enhance information transfer to others (“pushing”). Respondents in our study described an increased ability to share detailed patient information with colleagues and provide input to others’ care plans. However, when tools are unidirectional, it is not always possible for users to obtain sufficient information about a patient or plan or to confirm that information has been received [44,45]. The patient-centered messaging module was bidirectional and thus had the potential
to enhance information pushing when confirmation of receipt was required. It is also possible that once it is more widely used across disciplines, this module will increase the occurrence of electronic nonurgent information seeking (“pulling”).

Strengths and Limitations
One of the strengths of our study is its focus on the communication and collaboration experience of interprofessional teams (as opposed to documentation or other usual health record functions) that are critical to the care of patients with complex needs. Our team members were involved in both primary and secondary analyses, thereby improving rigor [46]. This study was also carried out in a real-world busy suburban community teaching hospital, and findings will thus likely be relevant in a wide range of teaching and nonteaching environments. Few studies have examined the PEOU and PU of electronic collaboration and teamwork tools using a qualitative methodology [47,48], with a focus on transferrable lessons. However, our study does have several limitations. First, the tool studied is a home-grown tool designed at our institution, and many of our findings may not be generalizable to other settings. We mitigated this limitation by conducting a secondary analysis informed by TAM that allowed us to explore lessons of technology acceptance that are likely generalizable to other settings. Second, our study was conducted at a single institution in the hospital environment. The value of improved communication and collaboration may be greater in the community or across care settings and organizations. Recognizing this limitation, our team is actively working on applying the lessons learned from the hospital environment across care settings. Third, despite purposive sampling, most participants self-reported that they were very comfortable with the technology. Finally, this paper focuses on the TAM assessing PEOU and PU and does not take into account the complex organizational, cultural, and environmental contexts, which undoubtedly affect the use of technology in the health care setting. Despite its simplicity, the TAM is one of the most widely used frameworks in predicting information technology adoption, and it has shown validity and reliability in effectively assessing technology acceptance [49,50]. However, the simplicity of this model also receives substantial criticisms in that it oversimplifies the complexity of the sociotechnical system [51] by focusing only on individual users’ perceptions, beliefs, and intentions. Alternate approaches that recognize the complexity of issues surrounding the implementation of health information technology are required [51]. We chose this approach to clearly present lessons that may be important in the design of future tools. Looking at the factors of TAM using a qualitative approach also surfaces many of the organizational and social components (eg, mix of paper and electronic charts, provider workflow, and interruptions). We point readers to our published mixed methods paper that contains the nontechnology factors that we encountered in this project [24].

Conclusions
Well-designed electronic tools that support the communication and collaboration needs of interprofessional teams are uncommon. To increase PEOU, future system designers should adhere to known usability principles relating to visual designs, consider the optimal training needs of users, ensure that information is succinct and organized, consider additional features that improve workflow, and remove logistical barriers if the system is embedded in a mix of paper and electronic systems. Users are likely to perceive the usefulness of these systems in their ability to increase efficiency, to support improved handovers, to allow communication and collaboration when face-to-face interactions are inefficient or impractical, and to promote shared understanding among team members. Owing to the collaborative nature of these tools, their actual use by all team members may impact the PU.

Acknowledgments
This research was funded by research grants from the Canadian Medical Protective Association and the Physician Services Incorporated Foundation.

Conflicts of Interest
The electronic tool was developed at Trillium Health Partners (a publicly funded health care organization), and all authors were staff at the organization during the study period.

References


43. Kuziemsky CE, Varpio L. Describing the clinical communication space through a model of common ground: 'you don't know what you don't know'. AMIA Annu Symp Proc 2010 Nov 13;2010:407-411 [FREE Full text] [Medline: 21347010]


Abbreviations

- **HIS**: health information system
- **PEOU**: perceived ease of use

https://humanfactors.jmir.org/2023/1/e39051

JHIM Hum Factors 2023 | vol. 10 | e39051 | p.794

(page number not for citation purposes)
**PU:** perceived usefulness

**TAM:** Technology Acceptance Model
The Priorities of End Users of Emergency Department Electronic Health Records: Modified Delphi Study

Matthew Yip\textsuperscript{1,2}, BMSc, MSc; Alun Ackery\textsuperscript{1,3}, MD, MSc; Trevor Jamieson\textsuperscript{1,3}, MD, MBI; Shaun Mehta\textsuperscript{1,3,4}, MD, GPLLM

\textsuperscript{1}Temerty Faculty of Medicine, University of Toronto, Toronto, ON, Canada
\textsuperscript{2}The Institute for Health Policy, Management, and Evaluation, University of Toronto, Toronto, ON, Canada
\textsuperscript{3}Unity Health Toronto, Toronto, ON, Canada
\textsuperscript{4}Department of Emergency Medicine, North York General Hospital, North York, ON, Canada

Corresponding Author:
Matthew Yip, BMSc, MSc
Temerty Faculty of Medicine
University of Toronto
1 King's College Circle
Toronto, ON, M5S 1A8
Canada
Phone: 1 6479092917
Email: matt.yip@mail.utoronto.ca

Abstract

Background: The needs of the emergency department (ED) pose unique challenges to modern electronic health record (EHR) systems. A diverse case load of high-acuity, high-complexity presentations, and ambulatory patients, all requiring multiple transitions of care, creates a rich environment through which to critically examine EHRs.

Objective: This investigation aims to capture and analyze the perspective of end users of EHR about the strengths, limitations, and future priorities for EHR in the setting of the ED.

Methods: In the first phase of this investigation, a literature search was conducted to identify 5 key usage categories of ED EHRs. Using key usage categories in the first phase, a modified Delphi study was conducted with a group of 12 panelists with expertise in both emergency medicine and health informatics. Across 3 rounds of surveys, panelists generated and refined a list of strengths, limitations, and key priorities.

Results: The findings from this investigation highlighted the preference of panelists for features maximizing functionality of basic clinical features relative to features of disruptive innovation.

Conclusions: By capturing the perspectives of end users in the ED, this investigation highlights areas for the improvement or development of future EHRs in acute care settings.

(JMIR Hum Factors 2023;10:e43103) doi:10.2196/43103

Keywords

Delphi; EHR; electronic health record; emergency medicine; emergency; functionality; health information exchange; health system; medical informatics; patient-physician relationship; usability

Introduction

Modern electronic health record (EHR) systems face difficulties meeting the unique needs of the emergency department (ED) \cite{1-3}. High volumes of patients through the ED drive documentation burden; high-acuity cases demand efficient deployment of care measures; diagnostic uncertainty increases the need for clinical decision support tools; and the interdisciplinary, collaborative environment drives a need for EHRs to support efficient transitions of care \cite{4}. In addition to these challenges, changes to the field of emergency medicine over the last several decades increase the need for highly efficient and capable information systems. As the complexity of patient’s presentations to the ED increases, measures of departmental crowding rise \cite{5}. Complexity and nuance to treatment plans further increase need to leverage digital health tools in the management of complex patients to improve clinical decision-making and patient outcomes, albeit with increasing
complexity of our digital systems [6-8]. The current COVID-19–mediated health human resource crisis has only exacerbated these challenges.

The International Standards Organization defines usability as “the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use” [9]. In the context of the ED, the specified goals of end users of an EHR may take on a variety of perspectives, given the different demands of this clinical space. A study evaluating the user-centered design principles of 11 EHR developers found that more than half of the developers had limited to inadequate interactions with clinicians in the development process of their products [10]. Despite the complexity of the unique needs of an ED EHR, there is a gap in the literature examining the perspective of the end user in an emergency medicine setting.

Delphi methods are a validated survey method to establish consensus opinion from a panel of experts [11]. The traditional Delphi process involves 3 rounds of information gathering: an initial round consisting of open-ended, qualitative questions followed by 2 rounds of Likert-scale rankings that allow for relative prioritization [11]. This process may be modified by introducing an initial set of parameters to narrow the scope of discussion [12-14]. A modified Delphi method offers the benefit of allowing focused discussion around specific attributes of a given problem. Delphi methods are unique in their ability to handle mixed types of information, both qualitative and quantitative in nature. They have previously been employed in the emergency medicine setting across several areas of investigation: investigating role definition of allied health team members, the development of violence screening criteria, the establishment of violence reduction strategies, and the selection of key performance indicators [15-18]. Delphi methods offer a validated method of synthesizing diverse perspectives about the current state and future improvements to ED EHRs.

To support hospital systems and practitioners develop future procurement criteria, and prioritize modifications, additions, or upgrades to their existing EHRs, we completed a systematic assessment of end user needs and priorities in the ED. This study aims to understand the nuances of perspective in physician end users regarding the ideal ED EHR.

**Methods**

**Identification of Key Usage Categories**

In phase 1 of our study, 2 independent reviewers completed review of academic literature on MEDLINE to build a list of usage categories of EHR. The reviewers also searched gray literature through web-based hand searches for topics related to information systems in acute care settings. After an iterative review of literature relating to both emergency medicine settings and EHR, 5 usage categories were developed inductively by the 2 reviewers. The findings were discussed with a working group comprised of 4 investigators with expertise in emergency medicine, health systems, and health informatics. The working group came to an agreement about 5 proposed key usage categories that were inputted into phase 2 to narrow the focus of discussion.

**Establishing Group Consensus Through Delphi Methods**

Phase 2 used Delphi methods that involved sequential rounds of survey and data dissemination to experts in both emergency medicine and information systems regarding their perspectives on each of the 5 usage categories. Recruitment of expert panelists was done through purposive sampling beginning with 4 investigators identifying candidates with expertise in both the clinical environment of the ED and health informatics at 6 tertiary- and quaternary-care centers across southwestern Ontario, including 3 level 1 trauma centers. Subsequently, the identified candidates were also invited to provide information on other potential informants. In total, 12 expert panelists were recruited across several hospital systems with extensive experience in both emergency medicine and health information systems. The panelists were spread across 3 separate disciplines (7 of 12 in emergency medicine, 3 of 12 in pediatrics emergency medicine, and 2 of 12 in general internal medicine). Several panelists held multiple leadership roles in their departments, with 4 of 12 acting as either chief or deputy chief, 7 of 12 acting as department lead across roles in quality and safety, virtual care, artificial intelligence and machine learning, and quality improvement. Several panelists also performed adjacent clinical duties with 4 of 12 serving as Trauma Team Leaders. Two panelists also fulfilled C-level positions at their respective hospital systems for roles in medical informatics. All panelists were associated with the University of Toronto in teaching and academic roles.

The Delphi study was conducted in 3 rounds of surveys [11]. Survey administration was conducted using the Research Electronic Data Capture (REDCap 12.0.29) tools hosted at the University of Toronto [19,20]. To reduce bias in both survey responses and response analysis, the identity of all panelists was kept anonymous through the Delphi rounds. Panelists and investigators were unaware of the identity of panelist’s responses and panelists were not aware of the identity of other members of the Delphi panel until the conclusion of the study. The analysis of outputs from each round was conducted by 2 independent reviewers and consensus was established before circulation of findings to panelists between each round.

The first-round survey involved qualitative information gathering through free-text responses. Free-text responses were analyzed using NVivo (NVivo Version 12). First, responses were coded deductively, using usage categories defined in Phase I of the study. Second, sentiment coding was performed by NVivo’s sentiment analysis with manual adjustment and necessary recoding based on consensus by the 2 independent reviewers. Outputs were circulated to panelists for review. The second-round survey gathered quantitative information on the perceived importance of first-round outputs using Likert scales and qualitative free-text responses about areas of disagreement from first-round responses. The quantitative outputs from the second-round survey were analyzed using Microsoft Excel (MSO Version 2205; Microsoft Inc) to generate descriptive statistics around measured variables and the qualitative outputs.
from the second-round survey were circulated to the panelists [21]. The third-round survey focused on establishing a ranked list of priorities based on the second-round outputs with the highest perceived importance resulting in a ranked list of priorities for each usage category.

**Ethical Considerations**

Phase 2 received the approval of the Research Ethics Board through the University of Toronto (protocol #00040996).

**Results**

In total, the perspectives captured by the expert panel spanned 6 separate hospital sites and 5 separate EHRs. Across all 3 rounds of survey, there was full retention of the original cohort of 12 expert panelists with no loss to follow-up between rounds. By using 5 key usage categories established by the working group members in phase 1 of the project (Table 1), the first round of surveys gathered free-form responses about the current needs of each category and generated a list of 10 features per usage category for a total of 50 features. Through the second-round survey, the panelists narrowed down the list to 25 features across key usage categories. Finally, in the third round of the survey, the panelists prioritized the top 5 features in each usage category relative to one another, for a total of 25 priorities (Textbox 1). Analysis of free-text responses produced statements of strengths and weaknesses for each category (Table 2). Several panelists raised ideas that may fall under the term of potential disruptive innovation, defined by Clayton Christensen as, “an innovation that makes things simpler and more affordable, and ‘technology’ is a way of combining inputs of materials, components, information, labor, and energy into outputs of greater value” [22]. Based on the priorities defined in Textbox 1 and the free-text responses by Table 2, possible features and innovations have been mapped to a typical journey through the ED, as a conceptualization of what an EHR may look like with these suggestions implemented (Figure 1).

**Table 1.** Usage categories defined by literature review.

<table>
<thead>
<tr>
<th>Usage category</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Information input</strong></td>
<td>The methods by which patient information is added or modified by care providers through multiple mediums [23-27]</td>
<td>Mobile device access, dictation support, and multidisciplinary access</td>
</tr>
<tr>
<td><strong>Digital health tools</strong></td>
<td>Features that augment or streamline the provision of care by providers [28-30]</td>
<td>Clinical decision support and computerized physician order entry</td>
</tr>
<tr>
<td><strong>Usability</strong></td>
<td>The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use [10,31-36]</td>
<td>Personalized dashboards, customizable quick picks within order sets, and inbox and task management</td>
</tr>
<tr>
<td><strong>Clinical workflow</strong></td>
<td>EHR features that impact patient flow through the ED [37-42]</td>
<td>Multidisciplinary communication and tools for communicating with external care providers after a visit to the emergency department</td>
</tr>
<tr>
<td><strong>Research and data analytics</strong></td>
<td>EHR features that allow for the ability to investigate research questions or conduct quality improvement studies [43-45]</td>
<td>Artificial intelligence and machine learning algorithms or adherence to interoperability standards</td>
</tr>
</tbody>
</table>

*EHR: electronic health record.*

---

https://humanfactors.jmir.org/2023/1/e43103

JMIR Hum Factors 2023 | vol. 10 | e43103 | p.798

(page number not for citation purposes)
**Textbox 1.** Key priorities defined by Delphi outputs (1-5 to indicate their priority with 1 being the highest and 5 being the lowest).

<table>
<thead>
<tr>
<th>Information input</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Support for multi-author documentation</td>
<td>2. Include the ability to input picture documentation</td>
</tr>
<tr>
<td>3. Integrate digital ambient scribes to expedite note taking</td>
<td>4. Enable quick picks or user favorites for easily accessed orders</td>
</tr>
<tr>
<td>5. Auto-populate fields with information that has already been given during the visit (i.e., triage assessment, consults from other services) or already available (i.e., past visits or community health record databases)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Digital health tools</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Streamlined governance structures to support pushing and pulling data from an electronic health record (EHR)</td>
<td>2. Integration of digital ambient scribes to expedite documentation time and order set suggestions</td>
</tr>
<tr>
<td>3. Identification of high-risk patients (i.e., poor prognosis and sepsis alerts)</td>
<td>4. Order entry and clinical decision support that builds on existing history for a given patient and continues to build on this history for subsequent visits</td>
</tr>
<tr>
<td>5. Embed clinical tools such as clinical practice guidelines or common risk stratification tools</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Usability</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Improve inbox and task management within EHR by allowing users to customize layout of their inbox</td>
<td>2. Streamline mobile access options that prioritize information input, similar to eCommerce or food delivery applications</td>
</tr>
<tr>
<td>3. Implementation of customizable home screen</td>
<td>4. Streamline access to other sources of information (i.e., community health record databases and previous medication reconciliations)</td>
</tr>
<tr>
<td>5. Streamline the number of required systems for different tasks or minimize disruption to workflow through improved integration</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Clinical workflow</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Support of patients beyond the hospital setting such as discharge instructions with prescriptions sent to an email or via SMS</td>
<td>2. Support for uploading documentation templates</td>
</tr>
<tr>
<td>3. Access imaging results within the EHR</td>
<td>4. Ability to communicate with others both inside the hospital setting (i.e., paging consults, porter services, and housekeeping) and beyond the hospital setting (i.e., community physicians, and emergency medical services)</td>
</tr>
<tr>
<td>5. Automatic data pulls from previous clinical documentation rather than manual chart review</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Research and data analytics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Improved governance structures that afford more flexibility to the end user with respect to access</td>
<td>2. Increase information access using role-based access (i.e., quality improvement lead, chief, and research roles), allowing for expedited data pulls and enabling queries for simple questions</td>
</tr>
<tr>
<td>3. Enhance standardization of coded information (i.e., diagnosis, chief complaints, and patient outcome) within sites and across sites</td>
<td>4. Embedded quality improvement tools</td>
</tr>
<tr>
<td>5. Embedded search engines to query and trend simple questions</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Strengths and limitations by category.

<table>
<thead>
<tr>
<th>Category</th>
<th>Strengths</th>
<th>Limitations</th>
</tr>
</thead>
</table>
| Information input         | • Improved accuracy of information in charting  
  • Improved collation of information and documentation for the overall care journey of a patient  
  • Support for verbal dictation methods expedites documentation | • Charting demands of EHRs\(^a\) increase documentation burden  
  • Redundancy of information input is attributed to the inability to carry over information previously gathered in the visit  
  • Some EHRs do not support all information formats (ie, pictures, ECGs) |
| Digital health tools      | • Order sets have increased the ease of use and safety is increased by decision support teams  
  • Current digital tools support patient safety  
  • EHRs have the technical capacity for deployment of innovative digital health tools, despite logistical difficulty and limitation of available health human resources | • Balancing innovative technology (ie, artificial intelligence, machine learning, and natural language processing) with patient safety, impossible to “try fast, fail fast” in the ED\(^b\) environment  
  • Governance structures such as privacy rules around information ownership, access rules within the department, limit accessibility of information |
| Usability                 | • Changes to order sets undergo a strict process to ensure that changes are in keeping with best available evidence  
  • Note templates are helpful in reducing documentation burden | • Standardization ensures patient safety but compromises flexibility of EHR  
  • “Look and feel” modifications are difficult to make with current systems  
  • Inbox and task management customization is not widely available |
| Clinical workflow         | • EHRs effectively collate information from past visits and current visit  
  • Makes interprofessional care between physicians, nurses, and clericals more seamless  
  • Data entered are more accessible and more legible | • Redundancy of gathering information and reinputting slows workflow  
  • Multiple systems are required for clinical tasks (ie, imaging results and past visits)  
  • Documentation burden reduces face-to-face time |
| Research and data analytics | • EHRs support data organization  
  • Increased ease of coding information in electronic form  
  • Supports a surplus of information relative to what is used | • Access to information is limited by privacy rules  
  • Steps of procedure for access to information for research is cumbersome, even for basic information or search queries  
  • Quality of information stored in the EHR due to lack of parametric data storage (ie, dropdown menus for diagnosis, checkboxes for signs, and symptoms) |

\(^a\)EHR: electronic health record.  
\(^b\)ED: emergency department.
Figure 1. A conceptualization of the intersection between the ranked priorities of panelists by usage categories and steps of the patient care journey. Innovations below each category are informed by Delphi outputs and offer a nonexhaustive view of proposed innovations across usage categories, affecting each step of care. Superscript after each priority denotes relative prioritization by panelists in Delphi rounds (1=highest priority, 5=lowest priority).

**Category I: Information Input**

Overall, it was found that panelists preferred that current EHRs improve on existing capabilities before trying to tackle potential disruptive innovations [22]. Panelists specifically listed and ranked digital ambient scribes, which process information from a patient–physician interview into a note in an attempt to reduce documentation burden, and auto-population of documentation from other sources of clinical information, lower than basic functionality such as multiauthor documentation and support for documentation of other forms of media. As strength, it was found that panelists thought that EHRs have streamlined the collation and standardization of information. A limitation of current information input capabilities of EHRs is the lack of support for multiauthor documentation, increasing the need to repeatedly gather, and document redundant information that has already been collected by other members of the patient’s care team. This drives documentation burden and creates inefficiencies.

**Category II: Digital Health Tools**

It was largely believed by panelists that human factors limit the implementation of digital health tools such as machine learning algorithms that provide clinical decision support, as opposed to the technical capacities of the current EHRs. Furthermore, the priorities list shows that panelists prioritize tools supporting clinicians in acute care settings such as identifying high-risk patients as opposed to pulling previous information from other sources such as previous charts or clinical portals. Panelists mostly expressed that EHRs have streamlined the ability to conduct repetitive, previously tedious tasks. However, they state that innovation requires large amounts of coordination and health human resources, so while the potential may exist for implementation, there may not be the current appetite or means to sustain this change.

**Category III: Usability**

The priorities of end users in this category saw 2 sentiments of thought, which first may seem conflicting. On the one hand, there was an interest in having increased customizability options within ED EHRs, such as the enablement of customization of quick picks and inbox management. However, there was also the argument for adaptation on the part of the end user to the features and limitations of the EHR. Overall, panelists believed that EHRs have increased standardization of care delivery through order sets that are vetted by central decision support teams, ensuring that orders are up to current care standards. However, in their current form, EHRs are limited in the customization options that they provide for their end users, even with respect to personal workflow features such as inbox task management, or “look and feel” customizations such as the layout of a given dashboard.

**Category IV: Clinical Workflow**

Panelists again prioritized basic functionality (ie, discharge planning, interdisciplinary communication) as opposed to disruptive innovation. Although EHRs have increased ease of collaboration among teams in the ED through collation of documentation from triage, panelists still raised concerns around the limitations of interoperability between hospital systems and other systems such as primary care EHRs. Additionally, even within a single-hospital system, it was found to be difficult to
communicate with other services that did not use the same EHR or charting method (ie, different clinical systems or paper charting).

**Category V: Research and Data Analytics**

Overall, panelists express that there was limitation with fluid access and usability of information. An undeniable strength of the EHR is that it has augmented the ability to collect, store, and access structured data. However, panelists identified that the ability to access the data in a meaningful way is still limited due to the format of stored data. Although it is possible to access volumes of information, the standardization of information input is lacking, such that any information sought for research purposes will still require manual recoding. Suggestions in this realm included improving drop-down menus to provide standardization of documentation input.

**Discussion**

**Principal Results**

The key usage categories developed in our investigation and the panelists’ priorities determined by Delphi outputs span several steps of a patient’s journey through the ED (Figure 1). These priorities highlight the balancing act that must occur in each usage category with the development and deployment of ED EHRs. With respect to information input, support for multiauthor documentation helps to reduce redundancy of information gathering and input, and support for innovations helps reduce documentation burden. With respect to digital health tools, improved governance structures could support the development and deployment of innovations that may aid in decision-making. With respect to usability, an optimized EHR for the ED would have customizability options for workflow and maintain strong standardization for deployment of care, such as order sets. With respect to clinical workflow support for communication beyond the hospital helps to ensure efficient and safe patient discharges, while consolidated information systems ensure efficient access to conducted investigations. With respect to research and data analytics, improved accessibility allows for more contribution from end users with respect to the development of new knowledge and useful clinical insights.

In Gawande’s [46] article titled, “Why Doctors Hate Their Computers,” Gawande writes of EHRs: “I’ve come to feel that a system that promised to increase my mastery over my work has, instead, increased my work’s mastery over me.” His assertion mainly centers around the collection of large amounts of unused information from a patient–physician encounter, which drives documentation burden and decreases patient–physician interaction time. Previous studies have estimated that the ED physicians may spend as much as 25% of the total time caring for a single patient on documentation [47]. Aligned with the previous literature and clinical experiences of documentation burden, Gawande highlights a key issue where EHRs can decrease efficiency and become a burden rather than a valuable tool.

These concerns are aligned with the findings from this investigation, with panelists broadly prioritizing functionality over disruptive innovation, and issues such as interoperability and the reduction of documentation burden being prioritized across several usage categories. For example, with respect to information input, support for multiauthor documentation and picture integration was prioritized over features such as digital ambient scribes or population from past documentation. Another example is seen in panelists’ priorities with respect to clinical workflow, where panelists prioritized discharge communication methods over auto-population of patient information from previous documentation. Panelists were sampled from a variety of care settings employing several different EHRs at each site, suggesting that no single EHR vendor comprehensively captures the priorities identified in this investigation.

By examining the discrepancies between the identified priorities of panelists and the qualitative responses of strengths and limitations, it is possible to identify areas for impactful improvements. For example, with respect to digital health tools, streamlined governance structures were identified as both a top priority (Textbox 1) and listed as a limitation (Table 2). Another usage category that demonstrated this was in research and data analytics, where panelists identified streamlined governance structures and increased role-based access as priorities (Textbox 1) and identified privacy as a limiting factor for gathering information (Table 2). Integrating this information identifies areas of high priority and can potentially inform prioritization of where system administrators can best optimize their own EHRs or build evidence-informed criteria in future acquisitions.

Compared to the deployment of Delphi methods in other emergency medicine clinical questions, the modifications to the process of this investigation optimized for depth of discussion in defined usage categories. The specific modification to the traditional process entailed defining the 5 usage categories through literature review which subsequently served as inputs to the Delphi model. Other investigations either integrate the literature review as one of the 3 traditional rounds or rely on free-text responses as a means to providing a focus of discussion [16,17]. A trade-off of the selected modification is that it prevents panelists from suggesting their own mental schema of usage categories of EHRs; however, this trade-off was made to achieve a deeper understanding of priorities within discrete categories. An additional benefit of a preliminary literature review is that focused discussion ensured concrete outputs from each round, which may have contributed to the complete retention of panelists across the 3 rounds of the Delphi process. Overall, through a preliminary literature review and a Delphi process with narrow targets based on prior inputs, the modified Delphi method strikes an appropriate balance between breadth and depth in the examination of ED EHRs.

**Limitations**

One potential limitation to this study is the generalizability of findings. Panelists are familiar with both ED care settings and health informatics in tertiary-care hospitals in southern Ontario, all with enterprise-wide deployments of their hospital EHR. This may lead to panelist-specific prioritization of other clinically adjacent activities such as academic research or data organization. Subspecialty interests may introduce additional variance to captured perspectives. Furthermore, this
investigation focused on capturing the perspectives of physicians as the end user, which does not capture the perspectives of other disciplines that engage with an ED EHR.

Conclusions

Improving EHRs to effectively meet the unique priorities of the ED demands a thorough understanding of the priorities of end users. A modified Delphi approach allows an in-depth analysis of perspectives of expert panelists in discretely defined usage categories. Capturing the perspectives of an expert panel from tertiary and quaternary care centers across Southwestern Ontario and served by diverse EHR vendors, the findings of this study highlight end-user prioritization of functionality over disruptive innovation. At a provider level, these findings will lead to meaningful reflection and discussions with department leadership about how an EHR can fit local needs. At an institution level, these findings will have implications for choosing future EHRs and adaptation of existing systems. At a developer level, these findings will have further sensitized developers to the preferences of end users in high-acuity settings. The future steps in discussions around EHR improvement should involve gathering the perspectives of allied health professionals who also engage with EHRs and with patients as they are the beneficiaries of improvements to information systems. Furthermore, comparison of perspectives gathered in the ED to perspectives from other areas of the hospital would establish commonalities, common pain points, and enhance our understanding of the information system preferences of end users.

Acknowledgments

The authors would like to thank expert panelists for their participation and contribution of their insights.

Data Availability

The data sets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Authors' Contributions

The authors confirm contribution to the paper as follows: research conception and design: MY, AA, TJ, and SM; project implementation: MY and SM; manuscript preparation: MY, AA, TJ, and SM. All authors reviewed and approved the final version of the manuscript. The authors received no financial support for the research, authorship, or publication of this paper.

Conflicts of Interest

None declared.

References


**Abbreviations**

- **ED**: emergency department
- **EHR**: electronic health record
- **REDCap**: Research Electronic Data Capture

Edited by A Kushniruk; submitted 29.09.22; peer-reviewed by KH Kim, C Slightham, D Chrimes; comments to author 27.11.22; revised version received 16.01.23; accepted 11.02.23; published 10.03.23.

Please cite as:

Yip M, Ackery A, Jamieson T, Mehta S

The Priorities of End Users of Emergency Department Electronic Health Records: Modified Delphi Study

JMIR Hum Factors 2023;10:e43103

URL: https://humanfactors.jmir.org/2023/1/e43103
doi: 10.2196/43103
PMID: 36897632
Performance of a Web-Based Reference Database With Natural Language Searching Capabilities: Usability Evaluation of DynaMed and Micromedex With Watson

Angela Rui1*, MA; Pamela M Garabedian2*, MS; Marlika Marceau2, BA; Ania Syrowatka1,3, PhD; Lynn A Volk2, MHS; Heba H Edrees1,3,4, PharmD; Diane L Seger2, RPh; Mary G Amato1,4, BCPS, MPH, PharmD; Jacob Cambre1, BS; Sevan Dulgarian1, BA, BS; Lisa P Newmark2, BA; Karen C Nanji2,3,5, MPH, MD; Petra Schultz6, BS, PharmD; Gretchen Purcell Jackson7,8, MD, PhD; Ronen Rozenblum1,3, MPH, PhD; David W Bates1,2,3,9, MS, MD

1Division of General Internal Medicine, Brigham and Women's Hospital, Boston, MA, United States
2Clinical and Quality Analysis, Mass General Brigham, Somerville, MA, United States
3Harvard Medical School, Boston, MA, United States
4Massachusetts College of Pharmacy and Health Sciences (MCPHS), Boston, MA, United States
5Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, Boston, MA, United States
6Merative, Cambridge, MA, United States
7Vanderbilt University Medical Center, Nashville, TN, United States
8Intuitive Surgical, Sunnyvale, CA, United States
9Harvard TH Chan School of Public Health, Boston, MA, United States
*these authors contributed equally

Corresponding Author:
Angela Rui, MA
Division of General Internal Medicine
Brigham and Women's Hospital
75 Francis St
Boston, MA, 02115
United States
Phone: 1 978 397 0082
Email: arui@partners.org

Related Article:
This is a corrected version. See correction statement: https://humanfactors.jmir.org/2023/1/e48468

Abstract

Background: Evidence-based point-of-care information (POCI) tools can facilitate patient safety and care by helping clinicians to answer disease state and drug information questions in less time and with less effort. However, these tools may also be visually challenging to navigate or lack the comprehensiveness needed to sufficiently address a medical issue.

Objective: This study aimed to collect clinicians’ feedback and directly observe their use of the combined POCI tool DynaMed and Micromedex with Watson, now known as DynaMedex. EBSCO partnered with IBM Watson Health, now known as Merative, to develop the combined tool as a resource for clinicians. We aimed to identify areas for refinement based on participant feedback and examine participant perceptions to inform further development.

Methods: Participants (N=43) within varying clinical roles and specialties were recruited from Brigham and Women’s Hospital and Massachusetts General Hospital in Boston, Massachusetts, United States, between August 10, 2021, and December 16, 2021, to take part in usability sessions aimed at evaluating the efficiency and effectiveness of, as well as satisfaction with, the DynaMed and Micromedex with Watson tool. Usability testing methods, including think aloud and observations of user behavior, were used to identify challenges regarding the combined tool. Data collection included measurements of time on task; task ease; satisfaction with the answer; posttest feedback on likes, dislikes, and perceived reliability of the tool; and interest in recommending the tool to a colleague.

https://humanfactors.jmir.org/2023/1/e48468
Results: On a 7-point Likert scale, pharmacists rated ease (mean 5.98, SD 1.38) and satisfaction (mean 6.31, SD 1.34) with the combined POCI tool higher than the physicians, nurse practitioner, and physician’s assistants (ease: mean 5.57, SD 1.64, and satisfaction: mean 5.82, SD 1.60). Pharmacists spent longer (mean 2 minutes, 26 seconds, SD 1 minute, 41 seconds) on average finding an answer to their question than the physicians, nurse practitioner, and physician’s assistants (mean 1 minute, 40 seconds, SD 1 minute, 23 seconds).

Conclusions: Overall, the tool performed well, but this usability evaluation identified multiple opportunities for improvement that would help inexperienced users.

(JMIR Hum Factors 2023;10:e43960) doi:10.2196/43960

KEYWORDS
medication safety; patient safety; usability; searching behavior; efficiency; quality of care; web-based databases; point-of-care information; POCI; point-of-care tools; artificial intelligence; machine learning; clinical decision support; natural language processing

Introduction

Background

Answering health care providers’ drug and disease questions in an accurate, effective, and efficient manner can be challenging. Common information-seeking issues that providers face include struggling to navigate through large amounts of information, being unaware of particular wording needed for optimal general search results, and requiring excessive time and effort to find answers [1-3]. Solutions may include the use of point-of-care information (POCI) tools, such as web-based databases that use evidence-based information to aid clinicians with drug and disease questions [4]. Commonly used drug and disease information systems that are considered POCI tools include UpToDate, DynaMed, Micromedex, and BMJ Best Practice [4,5]. POCI tools increase a provider’s ability to answer clinical questions in a timely manner, which can improve overall patient safety and care [4]. However, difficulties with searching in a manner that leads to a satisfactory answer may occur for various reasons, including the user not knowing when to stop searching or being unaware of how the POCI tool prefers clinical questions to be asked [6]. Artificial intelligence (AI) can be used to enhance POCI tools and has been integrated with electronic health records and clinical decision support systems to assist providers in improving patient and drug safety [7]. By combining AI capabilities, such as natural language processing (NLP), with the comprehensiveness of a POCI tool, there is potential to quickly answer a clinician’s questions and reduce mental fatigue compared with a manual search [8]. NLP is used in a variety of applications in health care and has been shown to assist in more efficient retrieval of information [9-11].

The DynaMed and Micromedex with Watson combined solution, now known as DynaMedex, is a POCI tool that includes drug and disease information with AI capabilities for information retrieval [12,13]. EBSCO partnered with IBM Watson Health, now known as Merative, to develop the combined tool, which aims to assist clinicians with answering clinical questions using evidence-based information [14,15]. This system combines the existing tools DynaMed, Micromedex, and Watson Assistant into an all-in-one web platform [16]. DynaMed is a medical condition knowledge database that contains summaries of evidence-based research, guidelines, clinical photographs, and other additional resources [17]. DynaMed provides peer-reviewed clinical content for 28 specialties on disease topics, health conditions, abnormal findings, disease evaluation, differential diagnosis, and disease management [16]. Micromedex is a pharmacological knowledge base with supporting literature curated for clinical significance by experts [18]. Micromedex is one of the largest web-based reference databases for medication information and provides detailed information on drug-drug interactions, drug monographs, and management of drug reactions [16,18]. Micromedex is often used by health systems to support the clinician in medication therapy management and patient education [8]. The purpose of combining DynaMed with Micromedex was to bring together drug and disease content into a single source that could be used to aid clinicians in making informed clinical decisions [12]. Watson Assistant is an AI-based conversational agent powered by IBM’s DeepQA supercomputer Watson, which aids users in information retrieval through a combination of NLP and machine learning [19]. Other research and applications of NLP in health care today focus on pulling important information from patient records to aid in decision-making, whereas Watson Assistant is a conversational agent that responds to user questions [9,11,20] related to drug information, drug interactions, and intravenous compatibility by mining databases of evidence-based information [8,16]. Drug information topics include drug classes, dosing, administration, medication safety, mechanism of action, and pharmacokinetics [7]. A prior study demonstrated Watson Assistant’s potential to answer clinician questions; a reported 80% of queries within Watson Assistant’s domain of knowledge were correctly classified by the conversational agent [7]. The paper also provides detailed information about the system architecture of Micromedex, including Watson Assistant.

Objective

The objective of this study was to collect clinicians’ feedback and directly observe their use of the combined tool to identify potential areas for improvement and assess participants’ perceptions to inform further development. Specifically, we focused on whether provider roles made a difference in their experience of using DynaMed and Micromedex with Watson. Little research exists on the user interaction and usability of these types of tools for health care providers. We evaluated the usability of the combined tool to determine how well users were able to reach their search goals with efficiency, effectiveness,
and satisfaction. We asked participants to test the combined tool by using both the general search function and Watson Assistant throughout the testing session to evaluate benefits and challenges arising from using either feature to search for information. From these findings, we summarized key themes that were observed or raised by providers in varying roles. In doing so, we generated recommendations for improving the clinician experience while using the tool.

**Methods**

**Overview**

This summative usability study collected data on how participants used DynaMed and Micromedex with Watson to complete information-searching tasks for a set of clinical scenarios. We report on quantitative usability metrics as well as describe observed differences in user experience among roles and experience with reference tools [21-23]. IBM provided a free subscription to the tool to conduct usability testing.

**Ethics Approval**

This research project was reviewed and approved by the Mass General Brigham institutional review board (2021P000139).

**Informed Consent**

Verbal informed consent was obtained from all individual participants included in the study. As part of recruitment, participants were informed in writing that their deidentified data from the audio and video recordings would be used for research.

**Recruitment**

Clinicians were recruited from inpatient and outpatient sites affiliated with 2 academic medical centers in Boston, Massachusetts, United States: Brigham and Women’s Hospital and Massachusetts General Hospital. From August 10, 2021, to December 16, 2021, recruitment emails were sent to physicians, pharmacists, registered nurses (RNs), nurse practitioners (NPs), and physician’s assistants (PAs) practicing in the following specialties: internal medicine, neurology, cardiology, oncology or hematology, infectious diseases, and endocrinology. The participant population was chosen based on the intended users of the tool. To achieve a sufficient sample size across clinicians in each role, general care RNs and pharmacists were also recruited. Clinicians were recruited using purposive and network sampling strategies. Participants were compensated for participation.

Before testing, participants were asked about their clinical role, years spent in practice, whether they practiced in an outpatient or inpatient setting, and whether they had prior experience using DynaMed and Micromedex with Watson. Participants were assigned a participant ID that was used on all study and data collection materials [21].

**Scenario and Script Development**

The study’s research pharmacists (HHE, DLS, and MGA) compiled a list of real-world questions supplied by clinical pharmacists from various specialty areas for both inpatient and outpatient settings. These questions were then evaluated and categorized by specialty and clinical area. The pharmacists determined whether each question could be answered accurately by DynaMed and Micromedex with Watson using the general search function or Watson Assistant. A set of questions that could be answered were selected and reworded into clinical scenarios for usability testing (Multimedia Appendix 1). Scenario question content was developed to be relevant to common situations according to specialty. Each usability session included 7 scenarios that covered a range of different clinical areas (Table 1).

**Table 1.** Assigned question categories for each script. The scripts included 7 scenarios covering these question categories.

<table>
<thead>
<tr>
<th>Question categories</th>
<th>Clinical specialty scripts</th>
<th>Nursing script</th>
<th>Pharmacy script</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adverse drug reaction or toxicity</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Disease</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drugs of choice or indication or therapeutic</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Dosing or kinetics</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Interaction (drug or herb or laboratory or disease)</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Monitoring or laboratory test</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Pregnancy or lactation or breastfeeding</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Drug administration</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Stability or compatibility</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>

*aRegistered nurses received 2 drug administration questions: either an inpatient or outpatient drug administration question depending on their primary work setting and a second drug administration question regardless of setting.

A total of 56 scenarios were created for usability testing. Eight unique scripts were created: cardiology, endocrinology, hematology or oncology, infectious diseases, internal medicine, neurology, nursing, and general pharmacy (Multimedia Appendix 1). For all scripts, the following question categories required the user to initiate their search using the Watson Assistant functionality: adverse drug reaction or toxicity, interaction (drug or herb or laboratory or disease), and pregnancy or lactation. All other question categories required the clinician to use the general search functionality to find the answer for the scenario. Physicians, the NP, and PAs were assigned a script based on their specialty. Pharmacists were
either assigned a general pharmacy script or a specialty script to have sufficient sample sizes for each script type. RNs were assigned a nursing script.

Pilot Testing
A pilot usability test was conducted to refine the testing procedure. The participating clinician was given a version of the internal medicine usability script that contained the scenarios and posttask questions. Scenarios that were confusing to the participant were reviewed and reworded to avoid misinterpretation.

Usability Testing Procedure
Each usability session was conducted remotely (ie, via Zoom [Zoom Video Communications, Inc]) to address both safety and scheduling concerns amid the COVID-19 pandemic. Participants were informed of the nature of the study as well as the scenario testing procedure and given the opportunity to ask questions related to testing. They were also informed of the moderator’s role as a neutral observer and the research assistant’s role in recording data. Verbal informed consent was obtained to record the audio and video of the Zoom session. Participants were asked a series of demographic questions and about their experience with reference databases for disease and medication management (Multimedia Appendix 1). Next, using the chat function in Zoom, the moderator (PMG) sent the participants the web address to access the DynaMed and Micromedex with Watson tool. Participants were asked to open the web page and begin screen sharing. The moderator provided no training on the tool but did ensure that the participants knew where to locate the general search function and Watson Assistant (Figure 1). Next, the moderator asked the participant to read each scenario aloud and search for the answer using either the general search function or Watson Assistant as detailed in the task. As the participant used the tool, they were encouraged to verbalize their thought processes, expectations for specific functionality, and reactions to elements in the tool. If the participant was able to find the answer, they informed the moderator of the answer and that they had completed the task. The participant could end the task at any time.

In situations where the participant encountered an unexpected usability issue that prevented them from moving forward with the current task or subsequent tasks, the moderator provided a prompt that assisted the participant in discovering why they were encountering the issue. These assists were not intended to help participants navigate content but rather were provided after multiple unsuccessful attempts to use a specific tool feature that was preventing them from accessing content; for example, assists were provided to participants who were not able to move forward with a task because they were unaware that it was necessary to clear filters on Watson Assistant at the beginning of each search to ensure that the conversational agent incorporated the correct keywords when searching for information.

When the participant finished the task, they were asked to respond to 2 posttask questions, administered through the polling feature in Zoom. The first posttask question required the participant to rate the ease of finding the answer on a 7-point Likert scale ranging from 1=very difficult to 7=very easy. The next posttask question asked the participant to rate their level of satisfaction with the answer using a 7-point Likert scale ranging from 1=very dissatisfied to 7=very satisfied. The participants were asked to explain their reasoning for each score. Finally, a semistructured posttest interview was conducted with the participants (Multimedia Appendix 1). Participants were able to provide their likes, dislikes, recommendations, and other opinions about their experience using the tool. The answers to the posttest interview questions were transcribed by the research assistant as the participant answered the questions.
Figure 1. Screenshot of DynaMed and Micromedex with Watson home page. Features include the general search bar and Watson Assistant. Since the time of the study, the user interface has been updated slightly (image courtesy of Merative, used with permission).

Analysis

During the usability testing sessions, data were logged into an Excel (Microsoft Windows 7; Microsoft Corp) spreadsheet and organized by participant identification and task number. The recordings of the usability sessions were analyzed by the reviewers (PMG, MM, JC, and SD) who were assigned to observe and record metrics pertaining to task success, time on task, and navigation and search behavior. The start time for each task was marked when the participant finished reading the scenario and asking any clarifying questions. The end time was marked when the participant found an answer or decided to end the task themselves. Technical issues or outside interruptions were removed from the total task time. To analyze navigation and search behavior, we captured the types and number of actions taken by the participants (eg, use of Find on Page or Ask Watson).

Announcements

ERICO Wins Frost & Sullivan Award

Our clinical decision products have been awarded the 2023 North America Product Leadership Award from Frost & Sullivan. Read the press release.

Monkeypox: The Latest

Read about monkeypox, including the 2022 outbreaks, clinical criteria for diagnosis, lesion progression (includes images), and more.

NEW COVID-19 Infection Control and Prevention

Open-Access Topic

Explore strategies, guidelines, and efficacy of infection control, immunization, prophylaxis, and screening measures for the prevention of COVID-19 in our COVID-19 Infection Control and Prevention topic.

Specialties

Browse and follow our evidence-driven and continuously updated specialty collections.
Ctrl-F, text entry into the general search function or Watson Assistant dialog box, or clicking on a search result or left navigation menu item). A content analysis was carried out on the posttest interview responses. Similar responses to each question were grouped and counted. All quantitative and qualitative data were compiled into a data set containing all metrics for all participants. The metrics, posttask question scores, and posttest interview answers were then analyzed by clinical question category and 3 role categories (physician, NP, and PA; pharmacist; and RN). We grouped the physicians, the solitary NP, and the PAs into 1 role category because they all work directly with patients to diagnose and treat health conditions. Descriptive statistics are reported for all quantitative metrics.

**Results**

**Participant Demographics**

Usability sessions were completed with 43 participants who had been practicing for an average of 10 years (Table 2). Of the 14 pharmacists, 5 (36%) received a specialty script (n=3, 60% had a specialty in cardiology and were assigned a cardiology script; n=1, 20% specialized in infectious diseases; and n=1, 20% specialized in hematology or oncology), and the remaining 9 (64%) received a general pharmacy script. Of the 2 RNs, 1 (50%) practiced in an inpatient setting, and 1 (50%) practiced in an outpatient setting. Of the remaining 27 clinicians, 21 (78%) were physicians, 1 (4%) was an NP, and 5 (19%) were PAs, and their specialties included cardiology (n=2, 7%), endocrinology (n=6, 22%), internal medicine (n=6, 22%), infectious disease (n=4, 15%), hematology or oncology (n=4, 15%), and neurology (n=5, 19%).

Most of the pharmacists (12/14, 86%) reported using Micromedex daily or at least once a week, whereas the physicians, NP, and PAs reported using Micromedex less than once a week in their current practice (22/27, 82%). Of the 21 physicians, 1 (5%) reported daily use of the combined solution on their mobile phone (Table 3).
<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clinical role</strong></td>
<td></td>
</tr>
<tr>
<td>Physician</td>
<td>21 (49)</td>
</tr>
<tr>
<td>Nurse practitioner</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Physician’s assistant</td>
<td>5 (12)</td>
</tr>
<tr>
<td>Pharmacist</td>
<td>14 (33)</td>
</tr>
<tr>
<td>Registered nurse</td>
<td>2 (5)</td>
</tr>
<tr>
<td><strong>Specialty used to assign script</strong></td>
<td></td>
</tr>
<tr>
<td>Cardiology</td>
<td></td>
</tr>
<tr>
<td>Physician</td>
<td>2 (5)</td>
</tr>
<tr>
<td>Pharmacist</td>
<td>3 (7)</td>
</tr>
<tr>
<td>Endocrinology</td>
<td></td>
</tr>
<tr>
<td>Physician</td>
<td>5 (12)</td>
</tr>
<tr>
<td>Nurse practitioner</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Internal medicine</td>
<td></td>
</tr>
<tr>
<td>Physician</td>
<td>5 (12)</td>
</tr>
<tr>
<td>Physician’s assistant</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Infectious diseases</td>
<td></td>
</tr>
<tr>
<td>Physician</td>
<td>4 (9)</td>
</tr>
<tr>
<td>Pharmacist</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Hematology or oncology</td>
<td></td>
</tr>
<tr>
<td>Physician</td>
<td>2 (5)</td>
</tr>
<tr>
<td>Physician’s assistant</td>
<td>2 (5)</td>
</tr>
<tr>
<td>Pharmacist</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Neurology</td>
<td></td>
</tr>
<tr>
<td>Physician</td>
<td>3 (7)</td>
</tr>
<tr>
<td>Physician’s assistant</td>
<td>2 (5)</td>
</tr>
<tr>
<td>Nursing: registered nurse</td>
<td>2 (5)</td>
</tr>
<tr>
<td>General pharmacy: pharmacist</td>
<td>9 (21)</td>
</tr>
<tr>
<td><strong>Years in practice</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;5</td>
<td>18 (42)</td>
</tr>
<tr>
<td>5-9</td>
<td>9 (21)</td>
</tr>
<tr>
<td>10-14</td>
<td>6 (14)</td>
</tr>
<tr>
<td>15-19</td>
<td>5 (12)</td>
</tr>
<tr>
<td>≥20</td>
<td>5 (12)</td>
</tr>
<tr>
<td><strong>Hospital setting</strong></td>
<td></td>
</tr>
<tr>
<td>Outpatient</td>
<td>15 (35)</td>
</tr>
<tr>
<td>Inpatient</td>
<td>20 (47)</td>
</tr>
<tr>
<td>Both outpatient and inpatient</td>
<td>8 (19)</td>
</tr>
</tbody>
</table>
Table 3. Reported frequency of Micromedex use (N=43).

<table>
<thead>
<tr>
<th>Role</th>
<th>Daily, n (%)</th>
<th>Once a week, n (%)</th>
<th>Once a month, n (%)</th>
<th>Few times a year, n (%)</th>
<th>Never, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physicians, nurse practitioner, and physician’s assistants (n=27)</td>
<td>2 (7)</td>
<td>3 (11)</td>
<td>3 (11)</td>
<td>5 (19)</td>
<td>14 (52)</td>
</tr>
<tr>
<td>Pharmacists (n=14)</td>
<td>6 (43)</td>
<td>6 (43)</td>
<td>N/A</td>
<td>N/A</td>
<td>2 (14)</td>
</tr>
<tr>
<td>Registered nurses (n=2)</td>
<td>N/A</td>
<td>N/A</td>
<td>1 (50)</td>
<td>N/A</td>
<td>1 (50)</td>
</tr>
<tr>
<td>Total</td>
<td>8 (19)</td>
<td>9 (21)</td>
<td>4 (9)</td>
<td>5 (12)</td>
<td>17 (40)</td>
</tr>
</tbody>
</table>

aN/A: not applicable.

Ease of Finding and Satisfaction With the Answer

All participants (N=43) completed the 7 scenarios for a total of 301 tasks. A participant was unable to use Watson Assistant because there were technical issues; therefore, they completed all 7 scenarios with the general search function only. The overall average ease of finding an answer was 5.68 (SD 1.57) out of 7 (physicians, NP, and PAs: 5.57, SD 1.64; pharmacists: 5.98, SD 1.38; and RNs: 5.07, SD 1.69). In 71% (5/7) of the clinical question categories, the pharmacists rated the ease of finding the answer higher than the physicians, NP, and PAs (Table 4). The largest difference between pharmacist ratings of ease of finding the answer (mean 6.36, SD 1.34) and physician, NP, and PA ratings (mean 5.19, SD 1.92) was in the adverse drug reaction or toxicity category (Table 4).

Overall, average satisfaction with the answer was 5.97 (SD 1.54) out of 7 (physicians, NP, and PAs: 5.82, SD 1.6; pharmacists: 6.31, SD 1.34; and RNs: 5.57, SD 1.79; Table 5).

Table 4. Average ease of finding the answer by question category and role.

<table>
<thead>
<tr>
<th>Question category</th>
<th>Tasks, n</th>
<th>Overall average (SD)</th>
<th>Physicains, nurse practitioner, and physician’s assistants (n=27), average (SD)</th>
<th>Pharmacists (n=14), average (SD)</th>
<th>Registered nurses (n=2), average (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adverse drug reaction or toxicitya</td>
<td>43</td>
<td>5.56 (1.79)</td>
<td>5.19 (1.92)</td>
<td>6.36 (1.34)</td>
<td>5.00 (1.41)</td>
</tr>
<tr>
<td>Disease</td>
<td>32</td>
<td>5.66 (1.26)</td>
<td>5.56 (1.31)</td>
<td>6.2 (0.84)b</td>
<td>N/Ac</td>
</tr>
<tr>
<td>Drugs of choice or indication or therapeutic</td>
<td>41</td>
<td>5.63 (1.69)</td>
<td>5.70 (1.46)</td>
<td>5.50 (1.51)</td>
<td>N/A</td>
</tr>
<tr>
<td>Dosing or kinetics</td>
<td>41</td>
<td>5.15 (1.90)</td>
<td>4.89 (1.99)</td>
<td>5.64 (1.69)</td>
<td>N/A</td>
</tr>
<tr>
<td>Drug administration</td>
<td>4</td>
<td>6.25 (2.92)</td>
<td>N/A</td>
<td>N/A</td>
<td>6.25 (0.96)</td>
</tr>
<tr>
<td>Interactiona</td>
<td>43</td>
<td>5.65 (1.77)</td>
<td>5.81 (1.78)</td>
<td>5.79 (1.37)</td>
<td>2.50 (2.12)</td>
</tr>
<tr>
<td>Monitoring or laboratory testing</td>
<td>43</td>
<td>6.02 (1.54)</td>
<td>5.96 (1.56)</td>
<td>6.21 (1.63)</td>
<td>5.50 (0.71)</td>
</tr>
<tr>
<td>Pregnancy or lactationa</td>
<td>43</td>
<td>5.95 (1.23)</td>
<td>5.85 (1.23)</td>
<td>6.43 (0.94)</td>
<td>4.00 (1.41)</td>
</tr>
<tr>
<td>Stability and compatibility</td>
<td>11</td>
<td>5.82 (1.08)</td>
<td>N/A</td>
<td>5.78 (1.09)b</td>
<td>6.00 (1.41)</td>
</tr>
</tbody>
</table>

aParticipants were asked to initiate search using Watson Assistant for these scenarios. One participant was unable to use Watson Assistant because of technical issues and used the general search function.

bA total of 5 pharmacist participants received a specialty script rather than the general pharmacy script; therefore, the counts by question category differ; 5 pharmacists completed a scenario in the disease category, and 9 completed a scenario in the stability and compatibility category.

aN/A: not applicable.

The pharmacists gave a rating of >6 to all but 2 question categories (disease and dosing or kinetics). The average ratings by the physicians, NP, and PAs ranged from 5.33 (SD 1.69) for the disease category to 6.19 (SD 1.21) for the drugs of choice or indication or therapeutic questions. As with average ease, the question category with the largest difference in satisfaction rating between the physicians, NP, and PAs (average 5.89, SD 1.67) and the pharmacists (average 6.71, SD 0.83) was adverse drug reaction or toxicity (Table 5).

Responses to overall ease and satisfaction varied by specialty (Table 6). The infectious disease specialists (pharmacists: 1/14, 7%; and physicians: 4/21, 19%) rated the ease of finding an answer as 4.89 (SD 1.62) out of 7 in comparison with the cardiology specialists, who had the highest average ease rating (6.11, SD 1.51). Average satisfaction with an answer was rated the highest by the pharmacists (6.44, SD 1.24) using the general pharmacy script compared with the internal medicine specialists, who rated satisfaction with the answer as 5.31 (SD 1.88).
Table 5. Average satisfaction by question category and role.

<table>
<thead>
<tr>
<th>Question category</th>
<th>Tasks, n</th>
<th>Overall average (SD)</th>
<th>Physicians, nurse practitioner, and physician’s assistants (n=27), average (SD)</th>
<th>Pharmacists (n=14), average (SD)</th>
<th>Registered nurses (n=2), average (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adverse drug reaction or toxicity(^a)</td>
<td>43</td>
<td>6.19 (1.45)</td>
<td>5.89 (1.67)</td>
<td>6.71 (0.83)</td>
<td>6.50 (0.71)</td>
</tr>
<tr>
<td>Disease</td>
<td>32</td>
<td>5.38 (1.58)</td>
<td>5.33 (1.69)</td>
<td>5.60 (0.89)(^b)</td>
<td>N/A(^c)</td>
</tr>
<tr>
<td>Drugs of choice or indication or therapeutic</td>
<td>41</td>
<td>6.15 (1.42)</td>
<td>6.19 (1.21)</td>
<td>6.07 (1.82)</td>
<td>N/A</td>
</tr>
<tr>
<td>Dosing or kinetics</td>
<td>41</td>
<td>5.63 (1.80)</td>
<td>5.52 (1.95)</td>
<td>5.86 (1.51)</td>
<td>N/A</td>
</tr>
<tr>
<td>Drug administration</td>
<td>4</td>
<td>7.00 (3.13)</td>
<td>N/A</td>
<td>N/A</td>
<td>7.00 (0)</td>
</tr>
<tr>
<td>Interaction(^a)</td>
<td>43</td>
<td>5.95 (1.84)</td>
<td>5.96 (1.85)</td>
<td>6.43 (1.28)</td>
<td>2.50 (2.12)</td>
</tr>
<tr>
<td>Monitoring or laboratory testing</td>
<td>43</td>
<td>5.98 (1.57)</td>
<td>5.81 (1.59)</td>
<td>6.36 (1.60)</td>
<td>5.50 (0.71)</td>
</tr>
<tr>
<td>Pregnancy or lactation(^a)</td>
<td>43</td>
<td>6.16 (1.13)</td>
<td>6.04 (1.09)</td>
<td>6.43 (1.28)</td>
<td>6.00 (0)</td>
</tr>
<tr>
<td>Stability and compatibility</td>
<td>11</td>
<td>6.27 (1.27)</td>
<td>N/A</td>
<td>6.67 (0.71)(^b)</td>
<td>4.50 (2.12)</td>
</tr>
</tbody>
</table>

\(^a\)Participants were asked to initiate search using Watson Assistant for these scenarios. One participant was unable to use Watson Assistant because of technical issues and used the general search function.

\(^b\)A total of 5 pharmacist participants received a specialty script rather than the general pharmacy script; therefore, the counts by question category differ; 5 pharmacists completed a scenario in the disease category, and 9 completed a scenario in the stability and compatibility category.

\(^c\)N/A: not applicable.

Table 6. Average ease and average satisfaction by specialty script.

<table>
<thead>
<tr>
<th>Specialty scripts</th>
<th>Role, n (%)</th>
<th>Average ease (SD)</th>
<th>Average satisfaction (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiology (n=5)</td>
<td>Pharmacist 3 (60)</td>
<td>6.11 (1.51)</td>
<td>6.06 (1.53)</td>
</tr>
<tr>
<td></td>
<td>Physician 2 (40)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Endocrinology (n=6)</td>
<td>Physician 5 (83)</td>
<td>5.74 (1.71)</td>
<td>6.19 (1.47)</td>
</tr>
<tr>
<td></td>
<td>Nurse practitioner 1 (17)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hematology or oncology (n=5)</td>
<td>Pharmacist 1 (20)</td>
<td>5.71 (1.56)</td>
<td>6.11 (1.35)</td>
</tr>
<tr>
<td></td>
<td>Physician 4 (80)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infectious diseases (n=5)</td>
<td>Pharmacist 1 (20)</td>
<td>4.89 (1.62)</td>
<td>5.37 (1.67)</td>
</tr>
<tr>
<td></td>
<td>Physician 4 (80)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internal medicine (n=6)</td>
<td>Physician 6 (100)</td>
<td>5.38 (1.71)</td>
<td>5.31 (1.88)</td>
</tr>
<tr>
<td>Neurology (n=5)</td>
<td>Physician 5 (100)</td>
<td>5.91 (1.44)</td>
<td>6.14 (1.33)</td>
</tr>
<tr>
<td>Nursing (n=2)</td>
<td>Registered nurse 2 (100)</td>
<td>5.07 (1.69)</td>
<td>5.57 (1.79)</td>
</tr>
<tr>
<td>General pharmacy (n=9)</td>
<td>Pharmacist 9 (100)</td>
<td>6.02 (1.28)</td>
<td>6.44 (1.24)</td>
</tr>
</tbody>
</table>

Time on Task
The total average time to find an answer across all tasks was 1 minute, 57 seconds (SD 1 minute, 32 seconds; range 00 minutes, 15 seconds-11 minutes, 36 seconds). The pharmacists took longer to finish tasks for all the question categories. The greatest differences in average time between the pharmacists (4 minutes, 8 seconds, SD 4 minutes) and the physicians, NP, and PAs (1 minute, 33 seconds, SD 39 seconds) were in the disease question category and drugs of choice or indication or therapeutic
category (pharmacists: 3 minutes, 17 seconds, SD 1 minute, 43 seconds; and physicians, NP, and PAs: 1 minute, 32 seconds; Table 7).

<table>
<thead>
<tr>
<th>Table 7. Time on task by question category and role.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question category</td>
</tr>
<tr>
<td>-------------------</td>
</tr>
<tr>
<td>All participants</td>
</tr>
<tr>
<td>Adverse drug reaction or toxicity(^a)</td>
</tr>
<tr>
<td>Disease</td>
</tr>
<tr>
<td>Drugs of choice or indication or therapeutic</td>
</tr>
<tr>
<td>Dosing or kinetics</td>
</tr>
<tr>
<td>Drug administration</td>
</tr>
<tr>
<td>Interaction(^a)</td>
</tr>
<tr>
<td>Monitoring or laboratory testing</td>
</tr>
<tr>
<td>Pregnancy or lactation(^a)</td>
</tr>
<tr>
<td>Stability and compatibility</td>
</tr>
</tbody>
</table>

\(\text{\(a\)}\) Participants were asked to initiate search using Watson Assistant for these scenarios. One participant was unable to use Watson Assistant because of technical issues and used the general search function.

\(\text{\(b\)}\) A total of 5 pharmacist participants received a specialty script rather than the general pharmacy script; therefore, the counts by question category differ; 5 pharmacists completed a scenario in the disease category, and 9 completed a scenario in the stability and compatibility category.

\(\text{\(c\)}\) N/A: not applicable.

Search and Navigation Behavior

The pharmacists took more actions on average (5.14, SD 3.48) than the physicians, NP, and PAs (4.2, SD 3.44) to find the answer to their questions and used the Ctrl-F or Find on Page feature in 30% (29/98) of the scenarios versus 15.3% (29/189) of the scenarios for the physicians, NP, and PAs (Table 8). In addition, in 31% (25/81) of the scenarios completed by the physicians, NP, and PAs and in 36% (15/42) of the scenarios completed by the pharmacists, they switched from Watson Assistant to view content on the main pages to obtain additional detail or because they were unable to find a satisfactory answer in Watson Assistant. Multiple general search entries and multiple Watson Assistant entries per scenario occurred in 28.7% (31/108) and 27% (22/81) of the scenarios, respectively, for the physicians, NP, and PAs. The pharmacists entered multiple general search entries per scenario in 41% (23/56) of the tasks.
Table 8. Navigation and search behavior by role.

<table>
<thead>
<tr>
<th>Navigation and search actions</th>
<th>Physicians, nurse practitioner, and physician’s assistants</th>
<th>Pharmacists</th>
<th>Registered nurses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of tasks using <em>Find on Page</em> or Ctrl-F, n (%)</td>
<td>29\textsuperscript{a} (15.3)</td>
<td>29\textsuperscript{b} (29.6)</td>
<td>0\textsuperscript{c} (0)</td>
</tr>
<tr>
<td>Count of switch from Watson Assistant to main content, n (%)</td>
<td>25\textsuperscript{d} (30.9)</td>
<td>15\textsuperscript{e} (35.7)</td>
<td>4\textsuperscript{f} (66.7)</td>
</tr>
<tr>
<td>Count of multiple general search entries, n (%)</td>
<td>31\textsuperscript{f} (28.7)</td>
<td>23\textsuperscript{b} (41.1)</td>
<td>3\textsuperscript{d} (37.5)</td>
</tr>
<tr>
<td>Count of multiple entries in Watson Assistant, n (%)</td>
<td>22\textsuperscript{d} (27.2)</td>
<td>7\textsuperscript{c} (16.7)</td>
<td>3\textsuperscript{f} (50)</td>
</tr>
<tr>
<td>Count of actions (clicks on search result, navigation menu, or entered text), average (SD)</td>
<td>4.2 (3.44)</td>
<td>5.14 (3.48)</td>
<td>5.92 (3.40)</td>
</tr>
</tbody>
</table>

\textsuperscript{a}n=189.
\textsuperscript{b}n=98.
\textsuperscript{c}n=14.
\textsuperscript{d}n=81.
\textsuperscript{e}n=42.
\textsuperscript{f}n=6.
\textsuperscript{g}n=108.
\textsuperscript{h}n=56.
\textsuperscript{i}n=8.

Posttest Interview Responses

Overall, 81% (35/43) of the participants felt that the information provided was accurate and reliable. Of the 14 pharmacists, 8 (57%) preferred the combined solution over their current system, and 10 (71%) would recommend the solution to their colleagues, whereas of the 27 physicians, NP, and PAs, 9 (33%) would prefer the combined solution, and 13 (48%) would recommend it to their colleagues (Table 9).

Table 9. Posttest interview responses by role and Micromedex experience (N=43).

<table>
<thead>
<tr>
<th>Question and role</th>
<th>Yes, n (%)</th>
<th>No, n (%)</th>
<th>Maybe, n (%)</th>
<th>Not yet, n (%)</th>
<th>I don’t know, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Would you recommend this tool to your colleagues?</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physicians, nurse practitioner, and physician’s assistants (n=27)</td>
<td>13 (48)</td>
<td>8 (30)</td>
<td>1 (4)</td>
<td>2 (7)</td>
<td>3 (11)</td>
</tr>
<tr>
<td>Pharmacists (n=14)</td>
<td>10 (71)</td>
<td>1 (7)</td>
<td>1 (7)</td>
<td>2 (14)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Registered nurses (n=2)</td>
<td>2 (100)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>All roles</td>
<td>25 (58)</td>
<td>9 (21)</td>
<td>2 (5)</td>
<td>4 (9)</td>
<td>3 (7)</td>
</tr>
<tr>
<td><strong>Did you feel the information was accurate and reliable?</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N/A\textsuperscript{a}</td>
</tr>
<tr>
<td>Physicians, nurse practitioner, and physician’s assistants (n=27)</td>
<td>21 (78)</td>
<td>1 (4)</td>
<td>5 (19)</td>
<td>N/A\textsuperscript{a}</td>
<td>N/A</td>
</tr>
<tr>
<td>Pharmacists (n=14)</td>
<td>12 (86)</td>
<td>0 (0)</td>
<td>2 (14)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Registered nurses (n=2)</td>
<td>2 (100)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>All roles</td>
<td>35 (81)</td>
<td>1 (2)</td>
<td>7 (16)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Would you prefer using DynaMed and Micromedex with Watson?</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physicians, nurse practitioner, and physician’s assistants (n=27)</td>
<td>9 (33)</td>
<td>11 (41)</td>
<td>3 (11)</td>
<td>4 (15)</td>
<td>N/A</td>
</tr>
<tr>
<td>Pharmacists (n=14)</td>
<td>8 (57)</td>
<td>4 (29)</td>
<td>1 (7)</td>
<td>1 (7)</td>
<td>N/A</td>
</tr>
<tr>
<td>Registered nurses (n=2)</td>
<td>2 (100)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>N/A</td>
</tr>
<tr>
<td>All roles</td>
<td>19 (44)</td>
<td>15 (35)</td>
<td>4 (9)</td>
<td>5 (12)</td>
<td>N/A</td>
</tr>
</tbody>
</table>

\textsuperscript{a}N/A: not applicable.

Participant Observations and Feedback

Think aloud and observations of user behavior highlighted usability issues with the combined tool. Participants experienced challenges with both the general search function and Watson Assistant in their ability to provide an exact match for their search. Participants often encountered issues when a search term was misspelled, an acronym or abbreviation was used, an unrecognized synonym was used, or too many words were entered. Another category of usability issues was related to the formatting and organization of the content pages. While looking through the content, participants made suggestions for additional
features to help locate information on the page (more graphs, tables, embedded links, visual cues, and consistency in formatting). Other issues were specific to Watson Assistant, particularly related to the challenges that participants experienced with closing the dialog box and understanding the interaction with the clear button and search terms.

In response to the posttask interview questions, participants shared that they liked the integration of the condition-specific knowledge with the detailed drug information. They also liked the inclusion of guidelines and citations accessible through hyperlinks. In addition, participants liked the predictive text and contextual suggestions in the general search. The drug-drug interaction feature of Watson Assistant as well as Watson Assistant’s ability to prompt them in a way that would narrow in on an answer were frequently mentioned as something they liked.

Participants reported their reasons for feeling that the information was accurate and reliable. The reasons included the following: answers matched their prior knowledge or experience in clinical settings, the tool cited evidence-based research such as clinical trials and guidelines, the evidence did not seem to be influenced by drug companies, and the tool was comprehensive with in-depth answers.

Participants disliked that the general search function and Watson Assistant both required specific words or phrasing to return quick and relevant results. They also expressed their dislike of the dense text, describing the length of time required to find an answer. Participants reported feeling as though they were not able to find answers to questions requiring more subspecialty specific knowledge.

**Discussion**

**Principal Findings**

Web-based medical information resources with technologies to support search are common, but few studies have been conducted to assess or improve their usability or to evaluate the ability of such tools to answer questions. We evaluated user interactions with the DynaMed and Micromedex with Watson combined solution and identified strengths as well as potential opportunities for improvement. We found that there was considerable variability in the time spent on each task. One reason for this included participant differences in their approach to consuming information; some were more interested in looking at additional detail and references, whereas others were satisfied with a general answer. This behavior is consistent with a prior study where physicians reported that a barrier to using electronic resources was related to difficulty in knowing when to stop searching for an answer [6]. Time on task could also have been influenced by participants' prior use of POCI tools because a connection has been reported between participants' prior experience and time to answer as well as confidence and satisfaction with the answer found [4]. Thus, the level of prior experience with these types of tools may also explain some of the differences seen among participants in terms of task ease and satisfaction.

Overall, pharmacists found DynaMed and Micromedex with Watson easier to use than other provider groups and were more satisfied with the answers (Table 5). Pharmacists generally had more experience with Micromedex, which may account for some of these findings. Pharmacists who had more experience with Micromedex may have had the advantage of understanding how best to articulate clinical questions for entry into the general search function to obtain helpful answers. In addition, pharmacists spent more time on tasks and performed more actions to find their answers. This extra effort could reflect their clinical role, where pharmacists routinely use POCI tools to help answer drug information questions for other clinicians. In comparison, the physicians, NP, and PAs had a different experience in which they tended to finish tasks in <2 minutes, with average ease and satisfaction ratings lower than those of pharmacists. Internal medicine physicians were the least satisfied with their answers. Physicians have been known in practice to experience considerable time constraints and cognitive burden from computer use [24,25]. Issues observed and verbalized by physicians included challenges scanning the text and the organization and visual hierarchy of the content pages, as well as the desire to have more curated knowledge useful for clinical practice. These results indicate that different requirements may be necessary to meet the needs of clinical pharmacists compared with those of practicing physicians, NPs, and PAs. These factors may explain why more pharmacists would recommend the combined tool to their colleagues than the physicians, NP, and PAs. It could also explain why more pharmacists would prefer using the combined tool over what they currently use than the physicians, NP, and PAs.

However, some changes could make the combined solution more accessible to physicians and RNs. Search functions that support a wider variety of differences in clinicians’ formation of clinical questions for general searches would be advantageous. A research study showed that medical residents had difficulty creating “answerable questions” to produce a general search that would provide the most relevant evidence [6]. Watson Assistant would also benefit from being able to accommodate more natural language used by clinicians, such as abbreviations, acronyms, and synonyms. A previous study demonstrated that the Watson Assistant conversational agent was able to match the intent of 80% of the queries answerable by the conversational agent [7]. Our study was able to examine the user interaction of different clinical roles with the combined solution through a set of standard scenarios known to be answerable by DynaMed and Micromedex with Watson. We found that although participants were successful and often able to ultimately find an answer, more than half of the participants (24/42, 57%) had to enter search terms in Watson Assistant multiple times, and slightly more than two-thirds of the participants (29/42, 69%) viewed website content outside of Watson Assistant to obtain more information at least once. In addition, for many of the participants (28/43, 65%), particularly pharmacists, use of the Ctrl-F feature to find information on the content pages was key to their success. Research on the usability of these types of tools and observations of clinician use is sparse. We found only 1 closely related study focusing on a comparison of efficiency, satisfaction, and accuracy of 2 tools, DynaMed and UpToDate [4]. The study found that clinicians were more
satisfied and found their answer more quickly with UpToDate, although the accuracy of the answers was similar using both tools. The authors hypothesized that greater familiarity with UpToDate may have influenced these results [4]. Our study aimed to uncover potential usability issues or use patterns by role that could help explain differences in satisfaction and efficiency in tools such as these. Studying the more detailed interactions with the system could have important benefits for practicing clinicians if the issues identified are shared with users and addressed in product enhancements.

Limitations
This study includes several limitations. First, users may have reasoned that the usability issues they encountered were not related to the design or performance of the DynaMed and Micromedex with Watson combined tool but rather because of the users’ own unfamiliarity with the tool [26]. This explains situations where the users struggled to find the answer but still provided high ratings when responding to their posttask questions. Second, we conducted usability testing with clinicians who all practice at the same organization, which may affect generalization of our findings to other institutions using DynaMed and Micromedex with Watson. Clinicians’ information-seeking behaviors may differ according to a variety of factors such as age, experience with technology, medical experience, geographic location, and access to resources [27]. Testing a larger number of participants and additional subgroups of users might offer more validity for generalizing findings [28]. Thus, it is important to continue usability testing with a wide range of clinicians to ensure that all user experiences can be recognized and addressed.

Third, although we attempted to standardize scripts by clinical question category and type, the scripts still contained differences (eg, question difficulty and wording). These differences could have affected participants’ experiences in understanding and answering the questions. In addition, creating individual scripts for specific roles (RNs and pharmacists) as well as specific specialties resulted in some inconsistencies related to which script to use for specific participants, for example, specialty pharmacists were interviewed using a specialty script instead of the general pharmacy script. This was done to reach sufficient sample sizes for the specialties that were more difficult to recruit. However, this proved to be helpful in providing insight into whether outcomes were influenced by specialty. Fourth and last, our study was conducted during the COVID-19 pandemic resulting in challenges recruiting a larger number of participants for each specialty. This was especially true for interested RN participants whose availability was greatly reduced by the increased demand of resources needed for patient care. To address both safety and scheduling concerns, usability testing sessions were conducted remotely, which may have influenced the results in comparison with in-person testing [29]. Additional testing with clinicians in various specialties may provide meaningful insights into the types of questions relevant to their practice, interaction behavior, and workflow considerations for these specialties, which could help to identify and prioritize tailored improvements to the combined tool.

Conclusions
This study is one of the first to test the usability of the DynaMed and Micromedex with Watson combined solution, now known as Dynamedex. It is also one of the first studies to compare ease and satisfaction of answers to questions in various content categories and by clinician role. We found that although the application performed well overall, pharmacists were able to use it most effectively in finding answers, whereas physicians and RNs had more difficulty finding the information they needed. We identified multiple changes that could be made to the tool to improve its usability, especially for inexperienced users. Understanding the determinants of information-seeking behavior is key to aiding physicians with finding answers to drug and disease management questions at the point of care.

Acknowledgments
RR and DWB are co-senior authors and contributed equally.

The authors would like to acknowledge the following IBM staff who provided feedback on DynaMed and Micromedex with Watson usability testing during the Brigham and Women’s Hospital–IBM meetings: Karlis Draulis, Courtney VanHouten, and Bryan Bohanan. Bryan Bohanan also answered questions from the Brigham and Women’s Hospital team about the combined tool. Editorial services for the manuscript were provided by Mya Baca. The authors would also like to acknowledge that AS was supported by a Fellowship Award from the Canadian Institutes of Health Research. This work has been supported by IBM Watson Health, now known as Merative (Cambridge, Massachusetts, United States), which is not responsible for the content or recommendations made.

Authors’ Contributions
All authors contributed to the study conception; design; and acquisition, analysis, or interpretation of the data. AR, PMG, AS, LAV, DLS, GPJ, and DWB were responsible for study conception or design. PMG, HHE, DLS, and MGA developed the interview guides. MM, AS, SD, and LPN conducted participant recruitment. PMG acted as the interview moderator and had either AR or MM assisting with data collection during testing. PMG, MM, JC, and SD abstracted the data from interview recordings. Data analysis was performed by PMG, MM, and AR. The first draft of the manuscript was written by AR and PMG, with all authors reviewing the draft and providing critical feedback. All authors contributed to and approved the final manuscript.
Conflicts of Interest

AR, PMG, MM, AS, LAV, HHE, MGA, SD, LPN, RR, and DWB received salary support from a grant funded by IBM Watson Health. DWB also reports, outside the submitted work, grants and personal fees from EarlySense, personal fees from CDI Negev, equity from Valera Health, equity from Clew, personal fees from MDCClone, personal fees and equity from AESOP, personal fees and equity from Feelbetter, and equity from Guided Clinical Solutions. RR reports holding equity in Hospitech Respiration, which makes Airway Management Solutions. RR also receives equity from TRI-O, which makes a medical device for diabetic foot ulcers; AEGEY Health, which provides an automated artificial intelligence–based diagnostic screening solution for retinal imaging; RxE2, which integrates the practice of pharmacy into clinical trials; and OtheReality, which provides a virtual reality technology to boost empathy in health care, all of which is unrelated to this work. RR is also receiving research funding from Boston Scientific Corporation, Telem, and MedAware, all of which is unrelated to this work. KCN reports author royalties from UpToDate Inc, equity from Guided Clinical Solutions, consulting fees from NORC at the University of Chicago, and research grants from the Agency for Healthcare Research and Quality and the Doris Duke Charitable Foundation. PS was employed by IBM Watson Health, and is currently employed by Merative, and holds stock in IBM. GPJ was employed by IBM Watson Health and now is employed by Intuitive Surgical. GPJ’s compensation from both IBM and Intuitive Surgical includes salary and equity. GPJ also serves on an advisory board for EBSCO.

Multimedia Appendix 1
Internal medicine usability test script: moderator guide.
[DOCX File, 24 KB - humanfactors_v10i1e43960_app1.docx ]

References


26. Holden RJ. People or systems? To blame is human. The fix is to engineer. Prof Saf 2009 Dec;54(12):34-41 [FREE Full text] [Medline: 21694753]


**Abbreviations**

AI: artificial intelligence  
NLP: natural language processing  
NP: nurse practitioner  
PA: physician’s assistant  
POCI: point-of-care information  
RN: registered nurse
A Visual Analytic Tool (VIADS) to Assist the Hypothesis Generation Process in Clinical Research: Mixed Methods Usability Study

Xia Jing1, MD, PhD; Vimla L Patel2, PhD; James J Cimino3, MD; Jay H Shubrook4, DO; Yuchun Zhou5, PhD; Brooke N Draghi1, BSc; Mychell A Ernst1, BSc; Chang Liu6, PhD; Sonsoles De Lacalle7, MD, PhD

1Department of Public Health Sciences, College of Behavioral, Social and Health Sciences, Clemson University, Clemson, SC, United States
3Informatics Institute, School of Medicine, University of Alabama at Birmingham, Birmingham, AL, United States
4Primary Care Department, College of Osteopathic Medicine, Touro University, Vallejo, CA, United States
5Department of Educational Studies, The Patton College of Education, Ohio University, Athens, OH, United States
6Electrical Engineering and Computer Science, Russ College of Engineering and Technology, Ohio University, Athens, OH, United States
7Health Science Program, California State University Channel Islands, Camarillo, CA, United States

Corresponding Author:
Xia Jing, MD, PhD
Department of Public Health Sciences
College of Behavioral, Social and Health Sciences
Clemson University
519 Edwards Hall
Clemson, SC, 29634
United States
Phone: 1 8646563347
Email: xia.xjing@gmail.com

Abstract

Background: Visualization can be a powerful tool to comprehend data sets, especially when they can be represented via hierarchical structures. Enhanced comprehension can facilitate the development of scientific hypotheses. However, the inclusion of excessive data can make visualizations overwhelming.

Objective: We developed a visual interactive analytic tool for filtering and summarizing large health data sets coded with hierarchical terminologies (VIADS). In this study, we evaluated the usability of VIADS for visualizing data sets of patient diagnoses and procedures coded in the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM).

Methods: We used mixed methods in the study. A group of 12 clinical researchers participated in the generation of data-driven hypotheses using the same data sets and time frame (a 1-hour training session and a 2-hour study session) utilizing VIADS via the think-aloud protocol. The audio and screen activities were recorded remotely. A modified version of the System Usability Scale (SUS) survey and a brief survey with open-ended questions were administered after the study to assess the usability of VIADS and verify their intense usage experience with VIADS.

Results: The range of SUS scores was 37.5 to 87.5. The mean SUS score for VIADS was 71.88 (out of a possible 100, SD 14.62), and the median SUS was 75. The participants unanimously agreed that VIADS offers new perspectives on data sets (12/12, 100%), while 75% (8/12) agreed that VIADS facilitates understanding, presentation, and interpretation of underlying data sets. The comments on the utility of VIADS were positive and aligned well with the design objectives of VIADS. The answers to the open-ended questions in the modified SUS provided specific suggestions regarding potential improvements for VIADS, and the identified problems with usability were used to update the tool.

Conclusions: This usability study demonstrates that VIADS is a usable tool for analyzing secondary data sets with good average usability, good SUS score, and favorable utility. Currently, VIADS accepts data sets with hierarchical codes and their corresponding frequencies. Consequently, only specific types of use cases are supported by the analytical results. Participants agreed, however, that VIADS provides new perspectives on data sets and is relatively easy to use. The VIADS functionalities most appreciated by participants were the ability to filter, summarize, compare, and visualize data.
Introduction

Data visualization, especially when data sets can be represented via hierarchical structures of biomedical terminology, has unique and superior advantages for human comprehension over other data presentation formats, such as tables and text [1]. However, the size of a visualization matters, as too much information can still be overwhelming even in this format [2-4]. Therefore, visualization alone may not be adequate to facilitate human comprehension. Instead, visualizing optimal sizes and complexity provides the desired enhancement to human comprehension of the underlying data sets.

Our visual interactive analytic tool for filtering and summarizing large health data sets coded with hierarchical terminologies (VIADS) is a secondary data analysis tool capable of providing visualization, filtering, analysis, summation, and comparison of data sets derived from the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) [5]; the International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) [6]; or the National Library of Medicine’s list of Medical Subject Headings (MeSH) [7] and their usage frequencies [8,9]. With existing ICD-9-CM codes, including diagnosis and procedure codes, and the steadily accumulating ICD-10-CM codes, numerous institutions and practices have data sets that VIADS can utilize. Meanwhile, PubMed continues to accumulate MeSH usage data, which VIADS can also use. By exploring summary views of underlying data sets or comparisons of similar data sets via VIADS, users can obtain overviews of data sets and highlights of the differences between the underlying data sets, which may aid in resource allocation decisions or comparisons of different but similar procedures or medications and their associated effects. In clinical research, the latter can facilitate hypothesis generation and validation. These are 2 typical VIADS use cases, one for health care administrators and the other for clinical researchers.

Our team developed the underlying algorithms and threshold settings for filtering and displaying such data sets using example applications. Additionally, we developed a free, publicly accessible web-based version of the tool for educational and research purposes [4,8-11]. Furthermore, VIADS can filter data sets by tuning thresholds to keep and present the most crucial data based on frequencies; visualizing results; comparing similar data sets (eg, data from 2005 versus 2015 or data between 2 hospitals); highlighting differences between data sets (ie, the most statistically significantly different ICD-9 codes between the 2 data sets); and summarizing results (ie, the aggregated results and displayed in the more generic and upper-level categories of the ICD-9 code system) using hierarchical terminologies, codes, and usage frequencies. VIADS could provide visualization (eg, the ICD-9 hierarchical structure, bar charts, and 3D plots) and interactive features (eg, when a user hovers the mouse on a node, more detailed information about that node in the data set will be provided; zoom in; various horizontal spacing layout options; select an algorithm and set thresholds accordingly) to assist users in determining thresholds when using VIADS to generate graphs. The comparative summary provided by VIADS compares 2 data sets. It displays the results in a single visualization, highlighting statistically significant differences (ie, ICD-9 codes) between the 2 data sets. Other research groups have recognized the unique value of visualizing hierarchical structures and have explored such relationships in medicine, social media, and information security [12-18].

In order to evaluate the usability and utility of VIADS, we designed and conducted a study to examine the process of generating clinical research hypotheses by clinical researchers with varying levels of experience (ie, the use case of VIADS by clinical researchers). This consisted of 2 groups of participants who used VIADS and 2 groups who did not. In each study session, all study groups used the same data sets (ie, ICD-9-CM diagnostic and procedural codes with frequencies) and the same time frame to generate data-driven hypotheses in the clinical research context [19]. The hypothesis generation process refers to the process researchers use to generate hypotheses. Some are data-driven, such as the process we used in the study session to generate hypotheses based on the data analysis results and visualization; others are observational-based, such as the unusual phenomena observed during wet lab experiments and the process between observing the phenomena and forming a hypothesis based on the phenomena.

The primary purposes of the study included the identification of (1) the potential role of VIADS in the generation of clinical research hypotheses, (2) the process of hypothesis generation in the context of clinical research, and (3) the role of experience level and its impact on the process of hypothesis generation. In this manuscript, we examine the usability of VIADS. We aimed to disseminate this VIADS usability study’s methods and findings to provide insight into the user interface design of secondary data analysis tools such as VIADS. We hope our experience will aid in the design and development of future data analysis software.

Methods

Methods for the Usability Study of VIADS

In this study, we used mixed methods. Participants in this study used VIADS for the hypothesis generation process. For this study, we modified the System Usability Scale (SUS; Multimedia Appendix 1) survey to assess the usability of VIADS. Brooke first proposed the SUS [20,21] in 1996, and it has been widely used to assess the usability of information systems for decades [22-25]. We modified the SUS by including...
open-ended questions that elaborate and clarify the Likert scale options. For example, if a user selected “disagree” or “strongly disagree” in response to the statement “I think VIADS is easy to use,” a follow-up question asked, “Can you please give an example of how VIADS is not easy to use?” This provided more specific feedback and determined why responses to specific items were unfavorable. The primary objective of this evaluation was to identify improvement opportunities for VIADS. Without explaining the respondent’s score selection, the SUS scores, in our opinion, lacked significant meaning. After the SUS evaluation, VIADS could be enhanced if some negative feedback could be addressed. As a result, we modified the standard SUS (ie, the follow-up questions can assist us in identifying areas that require improvement). Only negative responses were accompanied by a request for clarification.

**Utility Component of VIADS**

We administered a 6-question follow-up survey at the end of the study to verify the VIADS usage experience with possible responses of “yes,” “maybe,” “no,” and “Please elaborate on your answers” (open-ended, optional). Of the 6 questions, 1 question pertained to the overall usefulness of VIADS in clinical research, while the remaining 5 pertained to the specific ways in which VIADS could contribute to the research process. These questions focused on their perception of capacity to (1) provide novel perspectives, (2) facilitate data presentation, (3) facilitate results interpretation, (4) facilitate decision-making, and (5) facilitate other aspects of research. These questions are primarily aligned with the VIADS functionality, with the VIADS design objectives. These are subjective VIADS utility measurements; however, the answers are based on their 1-hour training and 2-hour intense use of VIADS. The objective measures of the utility of VIADS, such as a comparison of the quality of hypotheses generated via VIADS and without VIADS, are currently ongoing and will be shared with readers in separate manuscripts. The cognitive process analysis of the recorded think-aloud sessions is ongoing and will be published separately.

This usability evaluation study was conducted while the participants implemented the think-aloud technique with identical data sets to generate data-driven hypotheses using VIADS. All participants in the study adhered to the same protocol (Multimedia Appendix 2). Multimedia Appendix 3 contains the data extracted from the National Ambulatory Medical Care Survey (NAMCS) conducted by the Centers for Disease Control and Prevention [26,27]. We used data collected in 2005 and 2015 and preprocessed the NAMCS data sets by calculating and aggregating the ICD-9-CM diagnostic and procedural codes and their frequencies. VIADS accepts files in CSV format with 2 columns, one containing ICD-9 codes and the other containing the aggregated ICD-9 code frequencies. The same researcher conducted each study session remotely (via WebEx video conference).

Each participant had a 1-hour training session (Multimedia Appendix 4 contains the training slides that outline the primary functionalities and algorithms of VIADS) followed by a 2-hour study session. In each study session, a participant used the same data sets to perform the analysis; based on his or her experience and knowledge as well as the analysis results, hypotheses were generated, were recorded, and are currently being evaluated by an expert panel. An example of data analysis would be to examine the most frequently used ICD-9 codes in 1 year (2005 or 2015) or to compare the change in ICD-9 code frequencies between 2005 and 2015. During the study session, however, no particular algorithms were requested; each participant was free to explore any algorithms they desired. During the training sessions, the most commonly used scenarios of VIADS were demonstrated to each participant by the researcher. The results reported in this manuscript are based on the participants’ evaluations after the study sessions, which were recorded using BB FlashBack [28] to capture screen activities and conversations between each participant and the researcher. A professional transcription service subsequently transcribed the audio recordings. The modified SUS and an additional follow-up survey containing the 6 questions were administered after each study session. Participants were compensated based on their time spent on the study. Multimedia Appendix 5 is a VIADS user manual with additional information on how to use VIADS specifically.

The data-driven hypothesis generation process results are currently being encoded and analyzed. Once this step is complete, the results will be made public. Therefore, the quality of the hypotheses and the actual cognitive processes involved in hypothesis generation during each study session will be published separately.

**Ethics Approval**

The institutional review boards of Clemson University (IRB2020-056) and Ohio University (18-X-192) approved the study. All consent forms and study scripts were shared with all participants prior to the study sessions. The study data sets were shared with each participant on the day of the study session. Verbal permissions were obtained before the study sessions were recorded with each participant.

**Results**

**Overview of Results**

VIADS was tested by 12 participants, all clinical researchers. They were recruited through multiple national platforms, such as the American Medical Informatics Association discussion forums. Therefore, they were from geographically diverse institutions. Table 1 shows the demographic characteristics of the study participants.
Table 1. Demographic characteristics of participants in the usability evaluation of the visual interactive analytic tool for filtering and summarizing large health data sets coded with hierarchical terminologies (VIADS; n=12).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Results, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>5</td>
</tr>
<tr>
<td>Male</td>
<td>7</td>
</tr>
<tr>
<td>Age group (years)</td>
<td></td>
</tr>
<tr>
<td>&lt;35</td>
<td>6</td>
</tr>
<tr>
<td>35-45</td>
<td>2</td>
</tr>
<tr>
<td>46-55</td>
<td>4</td>
</tr>
<tr>
<td>Experience in clinical research (years)</td>
<td></td>
</tr>
<tr>
<td>&lt;2</td>
<td>6</td>
</tr>
<tr>
<td>2-5</td>
<td>3</td>
</tr>
<tr>
<td>5-10</td>
<td>3</td>
</tr>
<tr>
<td>Specialties</td>
<td></td>
</tr>
<tr>
<td>Health science</td>
<td>3</td>
</tr>
<tr>
<td>Internal medicine</td>
<td>3</td>
</tr>
<tr>
<td>Neurology</td>
<td>1</td>
</tr>
<tr>
<td>Pharmacy</td>
<td>2</td>
</tr>
<tr>
<td>Primary care</td>
<td>1</td>
</tr>
<tr>
<td>Other</td>
<td>2</td>
</tr>
</tbody>
</table>

SUS Results for VIADS

Table 2 shows the SUS scores for each participant. Among the 12 participants, 2 had SUS scores <60, and 5 had SUS scores ≥80. The scores ranged from 37.5 to 87.5. The overall mean SUS score for VIADS was 71.88 (SD 14.62), and the overall median SUS score was 75.

Table 3 presents the detailed raw SUS evaluation results for VIADS without SUS calculations. It summarizes the raw evaluation scores for each SUS evaluation item, with the following range of scores: strongly disagree=1 to strongly agree=5. For one-half of the questions in SUS, higher scores denoted more positive responses (direct questions); for the other one-half, lower scores indicated more positive responses (reverse questions).

The mean results for the direct questions ranged from 3.75 to 4.25 out of 5. The median score for all direct questions was 4. The scores for the reverse questions ranged from 1.92 to 2.83. For the reverse questions, 4 median scores were 2, and 1 median score was 3.

Table 2. System Usability Scale (SUS) scores for the visual interactive analytic tool for filtering and summarizing large health data sets coded with hierarchical terminologies (VIADS) from the individual participants (n=12).

<table>
<thead>
<tr>
<th>Participant number</th>
<th>SUS score</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>82.5</td>
</tr>
<tr>
<td>P2</td>
<td>85</td>
</tr>
<tr>
<td>P3</td>
<td>67.5</td>
</tr>
<tr>
<td>P4</td>
<td>72.5</td>
</tr>
<tr>
<td>P5</td>
<td>55</td>
</tr>
<tr>
<td>P6</td>
<td>65</td>
</tr>
<tr>
<td>P7</td>
<td>80</td>
</tr>
<tr>
<td>P8</td>
<td>85</td>
</tr>
<tr>
<td>P9</td>
<td>77.5</td>
</tr>
<tr>
<td>P10</td>
<td>37.5</td>
</tr>
<tr>
<td>P11</td>
<td>87.5</td>
</tr>
<tr>
<td>P12</td>
<td>67.5</td>
</tr>
</tbody>
</table>
Table 3. Detailed System Usability Scale (SUS) evaluation items and raw scores (n=12).

<table>
<thead>
<tr>
<th>SUS item a</th>
<th>Maximum score</th>
<th>Minimum score</th>
<th>Mean score</th>
<th>Median score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Would use frequently b</td>
<td>5</td>
<td>3</td>
<td>3.75</td>
<td>4</td>
</tr>
<tr>
<td>Unnecessarily complex c</td>
<td>4</td>
<td>1</td>
<td>2.33</td>
<td>2</td>
</tr>
<tr>
<td>Easy to use b</td>
<td>5</td>
<td>1</td>
<td>4.17</td>
<td>4</td>
</tr>
<tr>
<td>Need tech support to use c</td>
<td>4</td>
<td>1</td>
<td>2.50</td>
<td>2</td>
</tr>
<tr>
<td>Integrated well b</td>
<td>5</td>
<td>2</td>
<td>3.83</td>
<td>4</td>
</tr>
<tr>
<td>Inconsistencies c</td>
<td>3</td>
<td>1</td>
<td>1.92</td>
<td>2</td>
</tr>
<tr>
<td>Learned to use VIADS d quickly b</td>
<td>5</td>
<td>1</td>
<td>4.00</td>
<td>4</td>
</tr>
<tr>
<td>Cumbersome to use c</td>
<td>3</td>
<td>1</td>
<td>1.75</td>
<td>2</td>
</tr>
<tr>
<td>Can use confidently b</td>
<td>5</td>
<td>2</td>
<td>4.25</td>
<td>4</td>
</tr>
<tr>
<td>Need to learn more c</td>
<td>4</td>
<td>2</td>
<td>2.83</td>
<td>3</td>
</tr>
</tbody>
</table>

a Strongly disagree=1; strongly agree=5.
b Higher scores are favorable.
c Lower scores are favorable.
d VIADS: visual interactive analytic tool for filtering and summarizing large health data sets coded with hierarchical terminologies.

Utility Survey Results for VIADS

The modified SUS questionnaire and utility questions were asked and answered after a 1-hour training session and a 2-hour study session; when matched to the SUS scores, their answers corroborated their positive usage experience of VIADS. Table 4 presents the results of our VIADS utility questions. As indicated in Table 4, all results were separated into 3 categories: “Yes,” “Maybe,” or “No.” Among the respondents, 100% (12/12) agreed (ie, they all selected Yes) that VIADS provides new perspectives on the underlying data sets, 92% (11/12) felt that it could facilitate the presentation of data sets, and 75% (8/12) agreed that VIADS is a valuable tool for clinical research. Additionally, 75% (8/12) agreed that VIADS could facilitate the interpretation of results and decision-making in hypothesis generation. More than one-half (7/12, 58%) of the participants expressed conservative attitudes when asked if VIADS could assist with other aspects of research (ie, 58% selected either “maybe” or “no” as answers). In addition to subjective measures of the utility of VIADS, we published some objective measures at a conference [29]. For example, participants could generate 5 to 21 hypotheses within 2 hours, and the VIADS group took a shorter time, on average, to generate each hypothesis when we did not consider the quality of the hypotheses. More objective measures (such as the quality of the hypotheses) are still under analysis.

Table 4. The visual interactive analytic tool for filtering and summarizing large health data sets coded with hierarchical terminologies (VIADS) utility questions and results (n=12).

<table>
<thead>
<tr>
<th>VIADS utility survey item</th>
<th>Yes, n (%)</th>
<th>Maybe, n (%)</th>
<th>No, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provides new perspectives or measurements for data sets</td>
<td>12 (100)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Facilitates the interpretation of data sets</td>
<td>9 (75)</td>
<td>2 (17)</td>
<td>1 (8)</td>
</tr>
<tr>
<td>Facilitates decision-making in hypothesis generation</td>
<td>9 (75)</td>
<td>3 (25)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Facilitates the presentation of data sets</td>
<td>11 (92)</td>
<td>1 (8)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Useful in additional aspects of research</td>
<td>5 (42)</td>
<td>6 (50)</td>
<td>1 (8)</td>
</tr>
<tr>
<td>A useful tool for research overall</td>
<td>9 (75)</td>
<td>3 (25)</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

Qualitative Results From Open-ended Questions

Specific comments on answers to open-ended questions were organized as positive comments and suggestions, some of which were not positive. All positive comments were categorized under thematic headings, and only up to 3 items were presented in Table 5. The themes emerged after we aggregated and synthesized all comments from participants. The following insights for the improvement of VIADS were answers to the open-ended questions included in the modified SUS: (1) label data sets during comparison and carry the data set labels across pages, (2) more tips to explain the settings while uploading the data sets, (3) include the definitions of the terms and parameters used in VIADS, (4) the data sets accepted by VIADS are very specific, (5) provide further elaboration on the error messages, (6) provide a more detailed description of the functions.
Table 5. Thematic headings for the open-ended questions and examples for each theme.

<table>
<thead>
<tr>
<th>Thematic heading</th>
<th>Example statements</th>
</tr>
</thead>
<tbody>
<tr>
<td>VIADS facilitates the visualization of data sets to enhance understanding.</td>
<td>• “Pictorial and easy to read and understand huge data sets.”</td>
</tr>
<tr>
<td></td>
<td>• “VIADS presents a large data set containing diagnoses codes in an organized, intuitive graphical output with simple summary statistics that can be interpreted quickly and at a summary level, allowing for better understanding of the data set and how it can be analyzed.”</td>
</tr>
<tr>
<td></td>
<td>• “I think that VIADS would help with methodology, analysis, descriptive statistics, and presentation of results as well.”</td>
</tr>
<tr>
<td>VIADS provides a comparison function that compares similar data sets and highlights the results.</td>
<td>• “Comparison of complex data sets would be easy with this type of visualization.”</td>
</tr>
<tr>
<td></td>
<td>• “It is nice to have comparison of data sets, but that also goes back to understanding what the data set consists of.”</td>
</tr>
<tr>
<td></td>
<td>• “By comparing different sets of data, it would help clarify if it is an important/relevant area to study.”</td>
</tr>
<tr>
<td>The filtering function is a helpful means of reducing the size of data sets easily and effectively.</td>
<td>• “By being able to utilize large sets of data and recognize top percentages or number of certain topics, it helps you focus on an area to potentially study.”</td>
</tr>
<tr>
<td></td>
<td>• “I think the VIADS system is really cool, and it’s fascinating how it operates. I would say that it’s a great tool with algorithms in place but can also be overwhelming sometimes if there is overabundance of data. What’s great is that we can reduce the amount of data displayed. I think the fine tuning of threshold can be tricky though and difficult if we don’t understand what CC or NC or CC + ratio means.”</td>
</tr>
<tr>
<td>VIADS facilitates thought processes and hypothesis generation.</td>
<td>• “The many results and branches definitely help generate hypotheses. In the beginning, it is a little difficult since I would be focused on how to sort the data or minimize how many nodes/results are displayed, but after a while, with the key terms and diagnoses, it triggers my thought process so I think it could help with generating new ideas.”</td>
</tr>
<tr>
<td></td>
<td>• “VIADS answered many of the questions I had about the data set before using the tool. After using VIADS, I felt that some of my hypotheses would be valuable to pursue and some would not be as much. It also helped me build on some initial hypotheses to generate more specific and advanced questions.”</td>
</tr>
<tr>
<td></td>
<td>• “At least this session, the amount of diagnoses present and how it branches from one another helps not only stir up thoughts of known studies or information but helped me think of new ones or new questions that may not be answered yet.”</td>
</tr>
<tr>
<td>Other useful features of VIADS</td>
<td>• “I think that VIADS is very useful because it is simple and has a specific purpose that it serves well. I have used ICD codes frequently in my work so far, and I think this type of specific tool would be helpful for many applications.”</td>
</tr>
<tr>
<td></td>
<td>• “I think the training session helps a lot, because if I were to navigate it on my own, I probably wouldn’t know what the branching from each category means and the nodes when I did a preview, but it helps to click around.”</td>
</tr>
<tr>
<td></td>
<td>• “The graphical output and ability to export the diagram would be very useful in presenting diagnosis data.”</td>
</tr>
<tr>
<td>Suggestions</td>
<td>• “Not confident enough to base my hypothesis on numbers only.”</td>
</tr>
<tr>
<td></td>
<td>• “I think it could but then I would need training on how to create a data set to meet the VIADS system or how to set up complex data on a spreadsheet so when it’s uploaded to VIADS, it can help sort.”</td>
</tr>
<tr>
<td></td>
<td>• “Just that the buttons need to be adjusted...they don’t always ‘click’ unless you put your mouse on a very specific spot.”</td>
</tr>
</tbody>
</table>

aVIADS: visual interactive analytic tool for filtering and summarizing large health data sets coded with hierarchical terminologies.
bCC: class count.
cNC: node count.

**Discussion**

**Interpretation of the Results**

Previous research indicates that the mean SUS usability score is 68, on average, regardless of specific applications (eg, information systems or apps) [22]. The mean SUS score for VIADS in this study was 71.88, and the median score was 75. The literature shows that these are good usability scores [23,30]. Although the average score for VIADS can be improved further, it should be noted that VIADS is a complex analytic tool with many functionalities. The SUS score was encouraging, given the complexity of VIADS and participants’ heterogeneous backgrounds. Only 2 of the 12 participants had SUS scores <60. The rest had scores ≥65, and 5 had SUS scores ≥80. Table 2 includes the SUS score for each participant in the VIADS group.

Furthermore, the additional questions and constructive insights to improve the VIADS interface and instructions will help us to address these concerns more explicitly.

The average SUS score was 71.88, with an SD of 14.62, which is approximately 20% of the mean SUS. This large SD indicates heterogeneous opinions among participants about the usability of VIADS, allowing us to make more prudent and selective decisions about revisions to VIADS rather than implementing all suggestions. It is possible to investigate the variables contributing to such heterogeneity in a larger sample.

The feedback on the utility of VIADS was predominantly and consistently positive. The follow-up survey results provided some degree of the utility of VIADS, especially after 1 hour of training and 2 hours of using VIADS to analyze the data and generate hypotheses. As a secondary data analytic tool, VIADS
fulfills its design purposes. All participants agreed that VIADS offers new perspectives and measures of data sets. The usefulness of VIADS in facilitating data presentation (11/12, 92%), results interpretation, and decision-making in hypothesis generation was agreed upon by at least 75% of the participants. There appeared to be some reservations among the participants about making positive statements on additional aspects of clinical research beyond the dimensions about which they were explicitly asked. However, this could suggest that participants were only prepared to respond to items about which they felt sure. Therefore, we could take these results as additional validation of the positive nature of the overall results, acknowledging that there is always room for improvement.

Among all suggestions to improve VIADS among participants, suggestions 1 (ie, label data sets during comparison and carry the data set labels across pages), 2 (ie, more tips to explain the settings while uploading the data sets), and 5 (ie, provide further elaboration on the error messages) can be added to the VIADS interface. Suggestions 3 (ie, include the definitions of the terms and parameters used in VIADS) and 6 (ie, provide a more detailed description of the functions) are provided in the VIADS user manual and may be highlighted. There is also a legend key in the main interface. Furthermore, point 4 (ie, the data sets accepted by VIADS are very specific) is a limitation of VIADS; although the revisions are ongoing for all other points, point 4 has been excluded. To address point 4, a new tool is needed, which is under development.

Most of the participants positively commented on specific aspects of VIADS. However, it is possible that participants who provided lower SUS ratings were less inclined to leave comments on specific features.

In a system such as VIADS, it can be challenging to balance usability and utility. The functionality of the tool is not simple, and users must understand the underlying algorithms and how to use the tool’s various features and interpret the results it generates. The terms used in the interface alone (eg, NC is for node count, and CC is for the class count) represent a long list of definitions for users to grasp (Figure 1). The comparison summary of VIADS is presented using a single visualization (ie, the ICD-9 hierarchical structure), with highlighted ICD-9 codes if they are statistically different between the 2 data sets. During the development of VIADS, we devoted considerably more time to the utility of the tool, in terms of implementing the desired functionalities, than to the interface’s usability. Although we are encouraged by the SUS scores and the participants’ acclaim for VIADS’s primary features, which is their perception after their intense use of VIADS (ie, 1-hour training and 2 hours of use), actual performance measures are needed and ongoing.

Think-aloud protocols have been used as a method in the evaluation of information systems for decades. Some studies have focused on the investigations of the medical reasoning process [31-35], evaluation of clinical decision support systems [30,36,37], and additional purposes [25,38,39]. Our study used a think-aloud protocol to access the researchers’ thoughts, while participants used VIADS to assess its usability and utility.
Figure 1. (A) Screenshot of the visual interactive analytic tool for filtering and summarizing large health data sets coded with hierarchical terminologies (VIADS) showing the algorithms, generated graph, terms used in the interface, and definitions provided by VIADS; (B, C) enlarged portions of a graph generated by VIADS.

Significance of the Work

We asked the general research question: “Can secondary data analytic tools, such as VIADS, facilitate the hypothesis generation process?” One aspect of the tool related to this question is its usability. Thus, our objective was to investigate the tool’s usability and utility using mixed methods. The process of generating hypotheses using the same data sets via VIADS for clinical research projects was used as a task by participants, which provided real-use experience before participants answered the SUS and utility surveys. The results show the tool’s usability and some degree of utility. VIADS can be constantly updated with users’ feedback. This is an important first step to exploring the role of VIADS in facilitating clinical researchers to generate...
research and scientific hypotheses and support them at various levels of research.

Furthermore, this useful and accessible tool is freely available online as a user-friendly version, allowing users to leverage the tool without investing unnecessary time in technical details. Our research established a link between using a secondary data analysis tool and facilitating scientific hypothesis generation. This can be a starting point for utilizing secondary data analysis tools to understand the cognitive process of scientific hypothesis generation better.

Strengths and Limitations of This Study
The study included 12 participants, above the average range for a usability study. Past studies showed that 5 [38], 7 [30], 8 [36], and 12 [37] participants participated in comparable usability studies. The literature indicates that 5 participants can identify approximately 55% of usability issues, while 10 can identify approximately 80% [40]. With 12 participants, we are relatively confident that our usability study has a sufficient number of participants. In addition, our participants were selected from different regions of the country, with varying backgrounds within the clinical research context, providing a more comprehensive perspective of the tool.

Our SUS modification allowed participants to elaborate on the scores assigned to each SUS item. This allowed for targeted VIADS revisions. We believe that our modifications to the SUS were valuable and beneficial additions to the original SUS survey. Despite being grounded in the actual functionality of VIADS, the 6 utility questions and the SUS questions aligned well with the Health Information Technology Usability Evaluation Scale (Health-ITUES) [41]. In terms of health technology assessment frameworks [42,43], VIADS more closely resembles a data analysis tool than a mobile health application. Therefore, the economic evaluation of the tool’s impact deviates slightly from the tool’s primary purpose.

However, we know VIADS accepts only very specific types of data sets, not all. Consequently, the conclusions drawn from the data sets are specific rather than general. Now, we are developing a more generic supporting tool with a broader range of support for researchers.

Question 5 in Table 4 has the lowest agreeable rate; only 42% (5/12) of participants selected “Yes,” and 50% (6/12) selected “Maybe.” This question was supposed to capture any unintended impact of VIADS in addition to the 4 intended functionalities (ie, questions 1 to 4 in Table 4). However, the current presentation of the question can be confusing, which may lead to a low agreeable rate.

We recognize that our usability testing tool (SUS) captures the users’ perceptions, not how VIADS was used. Even though each participant had an intense VIADS use session before they completed the SUS survey, this still is a limitation of this study.

Due to lack of expertise, the graphs generated by VIADS consider more of the meanings and align with the underlying algorithms of VIADS, without much consideration of artistical aspects or color-blind users. Therefore, this is another limitation of this study. Even though there is no specific feedback on the artistic aspects of VIADS, this can be an area for improvement with appropriate additional expertise in the future.

Future Directions
We aim to increase the impact of VIADS through the (1) promotion of VIADS to increase its visibility among potential users and (2) development of new applications that facilitate the integration of VIADS with electronic health record systems or data repositories. This will enable VIADS to function as an add-on to existing systems that host large amounts of patient data. Through its analytical and visualization capabilities, the integrated version will streamline data sources, thereby promoting the adoption and use of the tool. Increasing the number of terminologies supported by VIADS is another possible area for further investigation. Finally, we could evaluate the tool at various stages and continuously use an iterative design process to improve VIADS.

Conclusion
VIADS, a tool that facilitates the generation of hypotheses in clinical research contexts, is a valuable addition to existing secondary data analysis tools. After intense use sessions, a diverse sample of clinical researchers perceived it to be useful and relatively usable. The new perspectives on hierarchical data sets and an easy-to-use interface provided by VIADS were recognized by users. The availability and use of ICD-9-CM, ICD-10-CM, and MeSH-coded data sets enable practical and convenient comparison of data sets and have many potential health care applications.

Acknowledgments
The authors express sincere thanks to all participants for their time and participation.

The project was supported by a grant from the National Library of Medicine of the United States National Institutes of Health (R15LM012941). It is partially supported by the National Institute of General Medical Sciences of the National Institutes of Health (P20 GM121342). The content is solely the authors’ responsibility and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Modified SUS survey (with utility questions) for evaluation of VIADS.
[PDF File (Adobe PDF File), 66 KB - humanfactors_v10i1e44644_app1.pdf]

Multimedia Appendix 2
Study script used for VIADS usability and utility study.
[PDF File (Adobe PDF File), 175 KB - humanfactors_v10i1e44644_app2.pdf]

Multimedia Appendix 3
Data sets used to conduct usability and utility study of VIADS.
[PDF File (Adobe PDF File), 345 KB - humanfactors_v10i1e44644_app3.pdf]

Multimedia Appendix 4
Training materials used during the training session for VIADS.
[PDF File (Adobe PDF File), 2990 KB - humanfactors_v10i1e44644_app4.pdf]

Multimedia Appendix 5
VIADS user manual.
[PDF File (Adobe PDF File), 102 KB - humanfactors_v10i1e44644_app5.pdf]

References


Abbreviations

- **Health-ITUES**: Health Information Technology Usability Evaluation Scale
- **ICD-9-CM**: International Classification of Diseases, Ninth Revision, Clinical Modification
- **ICD-10-CM**: International Classification of Diseases, Tenth Revision, Clinical Modification
- **MeSH**: Medical Subject Headings
- **NAMCS**: National Ambulatory Medical Care Survey
- **SUS**: System Usability Scale
- **VIADS**: Visual interactive analytic tool for filtering and summarizing large health data sets coded with hierarchical terminologies

© Xia Jing, Vimla L Patel, James J Cimino, Jay H Shubrook, Yuchun Zhou, Brooke N Draghi, Mytchell A Ernst, Chang Liu, Sonsoles De Lacalle. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 27.04.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Understanding the Role of Patient Portals in Fostering Interprofessional Collaboration Within Mental Health Care Settings: Mixed Methods Study

Keri Durocher1,2,3, MN; Hwayeon Danielle Shin1,4, MScN; Brian Lo1,4, MHI; Sheng Chen1, PhD; Clement Ma1, PhD; Gillian Strudwick1,4, PhD

1Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, ON, Canada
2Arthur Labatt Family School of Nursing, Western University, London, ON, Canada
3School of Health, Community Services, and Creative Design, Lambton College, Sarnia, ON, Canada
4Institute of Health Policy, Management, and Evaluation, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Keri Durocher, MN
Campbell Family Mental Health Research Institute
Centre for Addiction and Mental Health
60 White Squirrel Way
Toronto, ON, M6J 1H4
Canada
Phone: 1 (416) 599 1448
Email: kduroch3@uwo.ca

Abstract

Background: Patient portals are web-based systems through which patients can access their personal health information and communicate with their clinicians. The integration of patient portals into mental health care settings has been evolving over the past decade, as cumulated research to date has highlighted the potential role of portals in facilitating positive health outcomes. However, it is currently unknown whether portal use can foster interprofessional collaboration between clinicians and patients or whether the portal is a tool to support an already established collaborative relationship.

Objective: This mixed methods study aimed to understand how the use of a patient portal within mental health settings can impact the level of interprofessional collaboration between clinicians and patients.

Methods: This study was conducted in a large mental health care organization in Ontario, Canada. A convergent mixed methods design was used, where the primary data collection methods included questionnaires and semistructured interviews with patients who had experience using a portal for their mental health care. For the quantitative strand, participants completed the Health Care Communication Questionnaire and the Self-Empowerment subscale of the Mental Health Recovery Measure at 3 time points (baseline, 3 months of use, and 6 months of use) to measure changes in scores over time. For the qualitative strand, semistructured interviews were conducted at the 3-month time point to assess the elements of interprofessional collaboration associated with the portal.

Results: For the quantitative strand, 113 participants completed the questionnaire. For the Health Care Communication Questionnaire scores, the raw means of the total scores at the 3 time points were as follows: baseline, 43.01 (SD 7.28); three months, 43.19 (SD 6.65); and 6 months, 42.74 (SD 6.84). In the univariate model with time as the only independent variable, the scores did not differ significantly across the 3 time points (P=.70). For the Mental Health Recovery Measure scores, the raw mean total scores at the 3 time points were as follows: baseline, 10.77 (SD 3.63); three months, 11.09 (SD 3.81); and 6 months, 11.10 (SD 3.33). In the univariate model with time as the only independent variable, the scores did not differ significantly across the 3 time points (P=.34). For the qualitative strand, 10 participants were interviewed and identified various elements of how interprofessional collaboration can be supplemented through the use of a patient portal, including improved team functioning, communication, and conflict resolution.

Conclusions: Although the quantitative data produced nonsignificant findings in interprofessional collaboration scores over time, the patients’ narrative accounts described how the portal can support various interprofessional collaboration concepts, such as communication, leadership, and conflict resolution. This provides useful information for clinicians to support the interprofessional relationship when using a portal within a mental health setting.
Introduction

Patient Portals

A patient portal is a web-based system through which patients can access their personal health information and collaborate with their health care providers [1]. Around 2006, in North America, portals started to become widely adopted into health care settings through several private initiatives [2]. Since then, the integration of patient portals into various health disciplines and care areas has shown positive health care delivery outcomes, including improved quality of care and enhanced health status [3]. Specifically, positive health outcomes have been linked to care in the management of chronic diseases [4], cancer [5], and diabetes [6]. With such vast applications in multiple health contexts, research on portal applications within specific care areas, such as mental health care, is warranted.

Mental Health Care and Use of Portals

Over the last decade, research on patient portal integration into mental health care settings has evolved to build a body of knowledge on how to best support individuals within this broad health context. This research included various domains associated with the portal to understand the specific nuances of the evolving digital technology. Specifically, Etingen et al [7] performed a retrospective analysis through the Veterans Health Administration to determine whether individuals with specific diagnoses were more or less likely to access the portal. The researchers discovered that having anxiety disorders, posttraumatic stress disorder, and depression were associated with a greater likelihood of portal use [7]. Alternatively, Kipping et al [8] evaluated the benefits of implementing a portal for patients with mental illnesses [8]. Some noted benefits included a significant increase in appointment attendance and subjectively reported increases in autonomy [8]. Strudwick et al [9] studied various predictors of mental health professionals’ perceptions of using portals, such as their beliefs on whether patients should have portal access and whether they experience discomfort with this practice. Researchers discovered that perceptions of patient portal integration varied among different disciplines, such as psychiatrists reporting more negative perceptions of patient portals [9].

Interprofessional Collaboration

One notable gap in this evolving body of research is understanding how a patient portal can support interprofessional collaboration between mental health care providers, patients, and family members or caregivers. Although there is significant variation in the way interprofessional collaboration has been previously defined in the literature [10], we explored the concept of when ≥2 parties form a team, including clinicians, patients, and families or caregivers, and work concurrently to meet a desired outcome through shared power and partnerships [11]. Some of the key components of interprofessional collaboration include communication, role clarification, conflict resolution, leadership, team functioning, and patient-centered care, as defined by the Canadian Interprofessional Health Collaborative (CIHC) National Interprofessional Competency Framework [12]. This framework places the patient as central to the interprofessional collaborative partnership, which helps enhance joint decision-making [12]. Within mental health settings, previous research has linked treatment adherence with effective collaborative patient-clinician relationships [13]. However, it is currently unknown whether the use of a patient portal can enhance interprofessional collaboration between clinicians, patients, and families or caregivers or whether it is simply a tool to support an already established collaborative relationship.

Purpose and Research Objectives

The aim of this mixed methods study was to evaluate and understand the impact of patient portal use on the level of interprofessional collaboration from the perspectives of patients. The quantitative and qualitative objectives were as follows:

1. Quantitative strand: To determine whether the use of a patient portal has an impact on the level of interprofessional collaboration between patients and health care providers over time.

2. Qualitative strand: To understand how the use of a patient portal can influence patients’ perceptions of interprofessional collaboration with their health care providers.

Methods

Setting

This study was conducted within a large center that delivers mental health care in Toronto, Ontario, called The Centre for Addiction and Mental Health (CAMH). CAMH is the largest mental health teaching hospital in Canada, which delivers care to >34,000 patients per year across various inpatient and outpatient programs [14]. Globally, it is one of the world’s leading research institutions for mental health care.

Patient Portal

Several versions of patient portals exist across hospitals with slightly different functionalities. There are many similarities between different portals, but MyCare has been customized to the needs of CAMH and is only being used at CAMH. Through MyCare (Cerner patient portal), patients can access personal health information in collaboration with their health care team members. Patients can access parts of their electronic health record such as demographics, laboratory results, and clinician-written notes (eg, admission or discharge and
assessments). Other features of the portal include a secure clinician-patient messaging system and the ability to view upcoming appointments. These portals were integrated to select the outpatient service settings.

Design
This research is a part of a larger study that was completed over a 2-year period, and a protocol was previously published [15]. The primary data collection methods included questionnaires and semistructured interviews with patients and family members who had experience using a portal for their mental health care implemented at the organization.

This study included a secondary analysis of previously collected interview transcripts for the larger study, with a focus on interprofessional collaboration between patients and clinicians when using the portal. For a more in-depth explanation of the methods, please refer to the original protocol [15]. Publication of the larger study is currently in progress. The data were integrated at the design level using a fixed, convergent mixed methods study design [16,17]. Qualitative and quantitative data were independently gathered during a similar time frame and then compared to gain a further understanding of the topic of interest and participants’ experiences [16,17].

Recruitment
The participants were recruited using various techniques. One strategy was the distribution of recruitment flyers within the pamphlet that described how to use the portal. Interested participants were then provided with a link after registering for the portal, where recruitment information for the study was presented at the end of the email. Alternatively, when potential participants signed into the portal, the same recruitment information was included on the home page. If individuals had questions about the study, a research team member was on site within the outpatient settings during peak hours. Potential participants were also able to speak to the research team member in a private setting after they were enrolled to use the portal.

Sampling

Quantitative Strand
The minimum sample size for the quantitative strand was estimated to be 100 participants based on power calculation, as indicated in the study protocol [15]. All participants were assumed to be able to read English, as all components of the portal were available in English. Inclusion criteria were as follows: (1) being aged >16 years, (2) had enrolled to use the patient portal, and (3) self-reported having access to the portal for a time frame of <2 weeks. All participants were from outpatient clinical settings and provided written informed consent via REDCap (Research Electronic Data Capture; Vanderbilt University).

Qualitative Strand
At the end of the 3-month period for quantitative data collection, a convenient sample of participants was interviewed to discuss their experiences while using the portal. The inclusion criteria for the interviews included being a patient who had accessed and used the portal for at least 3 months. Participants also had to complete the quantitative questionnaires at baseline (before portal use) and after 3 months of use to be eligible for interviewing.

Data Collection

Quantitative Strand
All enrolled participants completed 2 questionnaires that encompassed crucial elements of interprofessional collaboration: the Health Care Communication Questionnaire (HCCQ) [18] and the Self-Empowerment subscale of the Mental Health Recovery Measure (MHRM) [19]. The questionnaires were administered via REDCap, a secure web application for collecting survey data. These questionnaires were administered at 3 time points: T0 (baseline), T1 (3 months of portal use), and T2 (6 months of portal use). Demographic information was also collected at the baseline data collection time point.

The HCCQ is a validated, 13-item scale that includes multiple elements of patients’ outpatient experiences, including problem-solving, respect, the lack of hostility, and nonverbal immediacy [18]. Each item refers to the concept of clinician communication, such as keeping calm, solving patient problems, and eye contact. Each item is scored on a 5-point Likert scale, with 0 meaning not at all to 5 meaning very much [18]. The maximum total score on the HCCQ is 65, with higher scores indicating more positive experiences of communication between patients and clinicians.

The MHRM is a 30-item self-report instrument [20], with all items being scored on a 5-point Likert scale (with a 0-4 range) for each associated item [19]. Overall scores for the MHRM can range from 0 to 120, with higher scores indicating higher levels of recovery-related experiences [21]. Self-empowerment is 1 of the 8 domains within this scale (items 5, 6, 7, and 8), and these 4 items (maximum score of 20) were analyzed in this study as a component of interprofessional collaboration.

Qualitative Strand
A semistructured interview guide was developed based on the objectives of the larger study. A total of 2 questions in the guide referred to interprofessional collaboration, and research assistants performed semistructured interviews using a secure videoconferencing platform (WebEx). The interviews were approximately 30 to 60 minutes in length and were completed between March 2021 and May 2022. With the consent of each participant, the interviews were audio recorded and transcribed verbatim. Any personal identifiers were removed from the transcripts before the data analysis was conducted.

Data Analysis

Quantitative Strand
All quantitative data analysis procedures were performed using SAS Enterprise Guide (version 7.15; SAS Institute). Participant characteristics at baseline (T0) were summarized using descriptive statistics. Continuous measures were summarized using means and SD, whereas categorical and ordinal measures were summarized using frequencies and proportions. Linear mixed effects models with random intercepts were used to model the trajectory of each outcome across the 3 study time points. Pairwise contrasts were generated between T0 and T1, between
T0 and T2, and between T1 and T2. The main analysis was not adjusted. We considered 2-sided \( P \) values <.05 as statistically significant.

**Qualitative Strand**

The CIHC National Interprofessional Competency Framework [12] is an established framework implemented for our secondary analysis of qualitative data. The interview transcripts were analyzed using a deductive approach [22], in which relevant domains of the framework were used. According to the CIHC, interprofessional collaboration involves active interprofessional relationships among team members, such as learners, health care professionals, and patients [12]. This CIHC framework has been implemented to study other phenomena, including interprofessional collaboration related to collaborative practice for providers [23] and advanced practice nursing [24]; however, it has not yet been applied to collaboration using a patient portal.

In total, 2 research team members (KD and HDS) were responsible for performing directed content thematic analysis of the interview data [22]. Both team members were registered nurses and PhD students with multiple years of experience in performing digital and mental health research. All transcription data were entered into NVivo Pro 11 (Lumivero) to facilitate coding and analysis procedures. As a pilot exercise, both team members coded 2 transcripts and reviewed any discrepancies before coding the rest of the data. After coding the remaining transcripts, collaborative thematic analysis was performed and mapped among 5 of the 6 themes within the framework.

**Integration**

To enhance our understanding of the quantitative and qualitative data, separate findings were reviewed simultaneously by the research team to understand how components of interprofessional collaboration relate to portal use. This helped the research team understand the contextual elements of how using the patient portal may relate to the elements of interprofessional collaboration among clinicians, patients, and family members or care partners.

**Ethics Approval and Informed Consent**

Ethics approval for the study was obtained from the Research Ethics Board at CAMH (REB 044/2018) and the University of Toronto (REB #40342). Written information about the study was provided to all potential participants, and an informed consent form was signed by all participants prior to being enrolled in the study.

**Results**

**Quantitative Strand**

**Demographic Characteristics**

A total of 113 participants were recruited for quantitative analysis. Of the 113 participants, 70 (62%) were aged between 26 and 64 years and 77 (68.1%) identified as female (Table 1). The most common diagnosis was a mood disorder, with 38.1% (43/113) of the participants reporting this. Regarding portal access, 99.1% (112/113) of the participants reported that they had daily access to the internet. Finally, on a scale of 0 to 100, participants reported their level of family or caregiver support, with a mean of 57.5 (SD 31.3).
Table 1. Demographic and clinical characteristics of participants (N=113).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Participant, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>77 (68.1)</td>
</tr>
<tr>
<td>Male</td>
<td>33 (29.2)</td>
</tr>
<tr>
<td>Prefer not to say</td>
<td>3 (2.7)</td>
</tr>
<tr>
<td><strong>Age range (years)</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;25</td>
<td>39 (34.5)</td>
</tr>
<tr>
<td>26-64</td>
<td>70 (61)</td>
</tr>
<tr>
<td>≥65</td>
<td>4 (3.5)</td>
</tr>
<tr>
<td><strong>Marital status</strong></td>
<td></td>
</tr>
<tr>
<td>Never married</td>
<td>62 (54.9)</td>
</tr>
<tr>
<td>Married, domestic partnership, common law</td>
<td>35 (31)</td>
</tr>
<tr>
<td>Widowed</td>
<td>3 (2.7)</td>
</tr>
<tr>
<td>Divorced or separated</td>
<td>11 (9.7)</td>
</tr>
<tr>
<td>Prefer not to answer</td>
<td>2 (1.8)</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
</tr>
<tr>
<td>Black or African American</td>
<td>8 (7.1)</td>
</tr>
<tr>
<td>East Asian</td>
<td>10 (8.9)</td>
</tr>
<tr>
<td>Hispanic or Latinx</td>
<td>3 (2.7)</td>
</tr>
<tr>
<td>Indigenous</td>
<td>2 (1.8)</td>
</tr>
<tr>
<td>South Asian</td>
<td>6 (5.3)</td>
</tr>
<tr>
<td>White or European</td>
<td>76 (67.3)</td>
</tr>
<tr>
<td>Mixed heritage</td>
<td>4 (3.5)</td>
</tr>
<tr>
<td>Other</td>
<td>4 (3.5)</td>
</tr>
<tr>
<td><strong>Diagnosis</strong></td>
<td></td>
</tr>
<tr>
<td>Anxiety</td>
<td>21 (18.6)</td>
</tr>
<tr>
<td>Mood disorder</td>
<td>43 (38.1)</td>
</tr>
<tr>
<td>Other</td>
<td>35 (31)</td>
</tr>
<tr>
<td>Prefer not to answer</td>
<td>9 (8)</td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>5 (4.4)</td>
</tr>
<tr>
<td><strong>Internet access</strong></td>
<td></td>
</tr>
<tr>
<td>Daily</td>
<td>112 (99.1)</td>
</tr>
<tr>
<td>Weekly</td>
<td>1 (0.9)</td>
</tr>
</tbody>
</table>

**Scales**

For both the HCCQ and MHRM-Self-Empowerment scales, from the original 113 participants who had T0 scores, 84 scores were recorded at T1 and 78 scores were recorded at T2. This is because for participants who had 1 more missing item, the total scores were not calculated for the descriptive analysis.

**HCCQ Score**

The mean total scores at the 3 time points were as follows: T0, 43.01 (SD 7.28); T1, 43.19 (SD 6.65); and T2, 42.74 (SD 6.84). In the univariate model with time as the only independent variable, the scores did not differ significantly across the 3 time points (P=.70). The estimated marginal means (least square means) were 42.96 (95% CI 41.60-44.33) for T0, 43.43 (95% CI 41.94-44.93) for T1, and 42.87 (95% CI 41.35-44.40) for T2 (Figure 1). Pairwise contrasts did not reveal substantial differences between T1 versus T0, T2 versus T0, and T1 versus T2. The HCCQ scores remained stable across the 3 time points.
**Self-Empowerment Scale**

The mean total scores at the 3 time points were as follows: T0, 10.77 (SD 3.63); T1, 11.09 (SD 3.81); and T2:11.10 (SD 3.33). In the univariate model with time as the only independent variable, the scores did not differ significantly across the 3 time points ($P=0.34$). The estimated marginal means (least square means) were 10.78 (95% CI 10.10-11.44) for T0, 11.23 (95% CI 10.50-11.96) for T1, and 11.07 (95% CI 10.32-11.82) for T2 (Figure 2). Pairwise comparisons of least squares means did not identify significant differences between T1 versus T0, T2 versus T0, and T1 versus T2.

**Figure 1.** Estimated marginal means of Health Care Communication Questionnaire (HCCQ) scores over time. Error bars denote 95% CIs.

**Figure 2.** Estimated marginal means of Mental Health Recovery Measure (MHRM) Self-Empowerment subscale scores over time. Error bars denote 95% CI.
Qualitative Strand

Overview

In total, 11 participants completed the interviews, and 1 interview was dropped during our analysis because there were no findings related to interprofessional collaboration. The following domains of the CIHC National Interprofessional Competency Framework were identified that applied to our analysis: Patient, Client, Family, Community-Centered Care; Team Functioning; Collaborative Leadership; Interprofessional Communication; and Interprofessional Conflict Resolution. Exemplar participant quotes of each theme are displayed narratively in the proceeding sections, with explanations of the integration of interprofessional collaboration with the patient portal in digital mental health care. For a full list of participant data mapped to the CIHC National Interprofessional Competency Framework, please see Multimedia Appendix 1.

Patient, Client, Family, Community-Centered Care

Integrating portals into mental health care settings can facilitate patient-centered care by enhancing the visibility of patients' pressing health needs. In addition, patients can review their clinical notes and verbally correct any misunderstandings or request clarification at the appointments with their clinicians. This process helps support the participation of patients and family members within the interprofessional circle of care and represents the core members of the care delivery pathway [12]. Two participants commented on how this process helped them meet their care needs and provided a sense of support and control:

I think for myself, I’m definitely the kind of person where I like being able to see the facts in front of me. I really like being able to have something written down, something concrete in front of my face, that helps me come to terms with things better, and be able to take the information and then work with it going forward. [Participant 9]

So, every so often when I was on the portal there would be these surveys like, how do you feel and how do you feel about your care? Those were great. I really liked getting them when I was in recovery. I felt like I was in control of my care a lot more than without the portal. [Participant 2]

Team Functioning

Functioning of the interprofessional team requires that all members have shared team dynamics that facilitate collaborative processes, including health care providers, patients, and their family members or care partners [12]. Through active participation, patients may feel a greater sense of control over their care outcomes. As the portal provides a channel for communication, collaborative goals can be understood by all team members, and patients may be more prepared for appointments:

I will go in more prepared with questions about...when I get a result when I’m there it’s kind of right away and you’re just trying to absorb it. So, I can check it at home, I can do my own little research, but then if I still have questions I can talk to the doctor and see what to do. [Participant 6]

It felt like I was kind of in control if that makes sense... It was nice to have that come so quickly because I’m so used to talking with a doctor and it takes like six weeks to hear back from my doctor. It kind of got rid of the anxiety of having to wait. There really was no wait and it was making me feel in control of everything. [Participant 7]

In addition, effective communication among the interprofessional care team can strengthen the working relationship among its members. When clinicians validate patients’ health needs and maintain ongoing communication through the use of the portal, this highlights its potential contribution to interprofessional collaboration. One participant commented on how the portal made them feel acknowledged even after their in-person appointment was finished:

I think, through the portal is kind of a way to acknowledge the fact that they are still paying it attention. They are still caring about your various health issues, whatever they may be. And it’s not like, once you leave the room, they forget about you. Not that that’s the case if you don’t have a portal, but it helps to solidify that, oh no, I am being acknowledged. My health is not being ignored, it’s right here, I’m seeing that they see it. [Participant 9]

Collaborative Leadership

To foster excellence in care, clinicians must include patients and family members or caregivers in a collaborative practice model [12]. In doing so, patients play a key role in their care responsibilities and can inquire about areas that must be clarified by clinicians. In addition, integration of the portal into practice can minimize the need for extra appointments for care areas that can be addressed through active portal use:

When I get a result when I’m there it’s kind of right away and you’re just trying to absorb it. So, I can check it at home, I can do my own little research, but then if I still have questions I can talk to the doctor and see what to do. [Participant 6]

I think it would reduce their need to do a lot of unnecessary paperwork. Let’s say they could write a prescription for some kind of drug and simply post it on the portal for the patient to print out and take to the drug store instead of, again, physically going to see the doctor, making the appointment, waiting in line, and doctors are always late. Basically, wasting a lot of everybody’s time just to get a piece of paper to take to a store when it can be accessed online. And the same thing, the instancy of information is a really great thing because it creates a good venue of communication between the patient and the doctor, not simply limited to the physical appointment. [Participant 4]

Collaborative leadership also shifts the responsibility of care to a joint approach between patients, families or caregivers, and clinicians when using a portal. Therefore, clinicians must be
cognizant of reducing the use of medical jargon to promote a digital environment for shared leadership [12] and ensure that patients are aware of certain medical terms and information:

I like when you talk to me as if I’m a colleague that you’re talking to and leave it to me to say to you, I don’t know what that term means, I don’t know what that definition is, and then you can backtrack and say, okay, so let me inform you. I prefer you not worrying about talking over me, as opposed to insulting me by talking down to me. That’s me. [Participant 5]

Interprofessional Communication

The process of interprofessional communication should include a collaborative and responsive approach between the clinician and patient, which can be supplemented using a portal. When patients can thoroughly understand care decisions, it can enhance the trusting relationship with their clinician. A total of 2 participants commented on how the portal provided this sense of trust and encouragement for having care discussions:

It gave me the chance to talk to them about some of the diagnoses. If they said, how do you feel about this diagnosis of bipolar rather than this other diagnosis of schizo affective, or whatever? It was good to know where that was coming from, and it was also good to know the reasoning behind it without having to waste time during a meeting with the psychiatrist or the doctor. [Participant 2]

I think a lot of people don’t trust their clinician, especially today because there’s a lot of misinformation out there and, I don’t know, people don’t always trust healthcare professionals. If you give someone access to the same information as a healthcare professional has access to then it, theoretically, would...It theoretically should increase the trust level there because I can... If I don’t think...I could look up that lab value. [Participant 1]

One participant also remarked how portal implementation can improve efficiencies in communication and reduce the need for having duplicate conversations:

That’s just a good record to have of what has been covered so that we don’t need to waste the appointment time, which is usually an hour or so, fairly short, on covering things that had already been covered. It’s good for that, I would say, and basically keeping track of the progress. So, seeing the whole transition from appointment to appointment and where that leads. [Participant 4]

Despite the positive aspects of how portals can enhance communication, 1 participant remarked that despite the integration of a patient portal, there may still be uncertainty regarding whether the clinician is fully forthcoming in what is placed within the portal for patient viewing. This demonstrates the importance of building a trusting, foundational relationship in addition to the implementation of supportive technology into care relationships:

So, it’s like I’m having the information relayed to me, like there’s a middleman, kind of. So, I think that there isn’t as much of a trust, necessarily. Or there’s always a bit of questioning of, well, am I getting the full story here? Am I getting the full scope of information that I need, or am I getting what they believe is all I need? So, being able to read it myself, I know that I’m being given the information because I’m seeing it in front of me. I know that what they said they do believe because they also included it for me to access. [Participant 9]

Interprofessional Conflict Resolution

When interprofessional relationships are developed between clinicians and patients, conflict can be an inevitable component of the ongoing caring relationship. As noted in the previous quote, uncertainty regarding what is included in clinical notes can be a potential source of conflict. However, working collaboratively to build consensus on issues and actively working to solve disagreements are strategies for conflict resolution that can be supplemented using a portal [12]. Some areas to consider are the portal design components and information entered by the clinician within the notes. In total, 2 participants commented on how these design elements are important, which speaks to missed opportunities for conflict resolution:

The comment that I had about the notes is it would have been nice for me to be able to flag certain things. I had been at an inpatient facility and one of the nurses there had given an account of events about how something had occurred. I would have really appreciated the opportunity to flag that and give my interpretation, because in the portal there was only one... it was great to see what was written, but there was only one side to it. [Participant 2]

Certainly, seeing doctors’ notes, what they said, may have been. Because I feel like I never really properly understood. If people paraphrase what I say, I find that they often change what I perceive is the meaning of my statement. So, if I could see someone writing notes and them not being accurate to the message I was trying to convey. [Participant 5]

One participant also remarked how the implementation of the portal could reduce some sources of conflict, such as questioning the usefulness of certain assessments or interventions. If patients can understand the rationale for these activities, this source of conflict can be reduced:

I was filling out these mood charts and then he would just file them away and I questioned if he was reading them, I questioned if I was wasting my time. I feel like maybe if I was submitting them on the portal, at least I’d feel like someone is looking at them in the meantime, like, I’m submitting them before I get there. [Participant 1]

Integration

Despite a lack of change on the scales related to communication and self-empowerment over time, participants revealed many...
different perceptions related to interprofessional collaboration through the semistructured interviews. This demonstrates how various elements of how interprofessional collaboration relate to portal use may be best described through the subjective, narrative experiences of patients. For example, a few participants commented on how interprofessional communication practices can improve through various components of the portal, such as through a preemptive chart review. Therefore, a notable benefit of this secondary analysis is that after merging these data, we now have a deeper understanding of some aspects of interprofessional collaboration that can be enhanced by portal use and other aspects that require further exploration.

Discussion

Principal Findings

The aim of this mixed methods study was to evaluate and understand the impact of portal use on patients’ experiences of interprofessional collaboration within a mental health context. Previous literature describes the role of technology in collaboration from the providers’ perspectives [25,26]. At the time of writing, this study is one of the first to assess the impact of portal use on collaboration from the patients’ perspectives, who are the central members of the care team. The quantitative results showed no significant findings, whereas the qualitative strand sheds light on the impact of portal use on multiple components of interprofessional collaboration beyond clinicians’ communication skills and patients’ sense of empowerment. For example, portals encouraged patients’ participation in their own care, promoting collaborative leadership and a sense of control. Furthermore, portals helped reshape traditional team dynamics, ensuring that patients are central members of the team, in contrast to research on interprofessional teams that primarily focuses on working interactions between different providers [27,28]. Therefore, the portal’s ability to encourage participation from patients is noteworthy because ensuring the full participation of patients as interprofessional collaborators can minimize professional paternalism [29,30]. Most notably, portal use does not seem to detract from promoting interprofessional collaboration.

Despite the potential for portal use in facilitating interprofessional collaboration in mental health care settings, there were a few areas of concern that must be acknowledged. For example, trusting relationships must be established between patients and health care providers. Otherwise, patients perceived that their notes are not fully disclosed to them, which could be a barrier to establishing true coleadership of patients and providers in the care team. In addition, interpreting clinical notes can be a challenge, which consistently have been reported in the current literature in mental health care settings [31] and beyond [32]. These gaps in the current practice of interprofessional collaboration when using a portal provide foundational criteria for building future directions in mental health settings.

Future Directions

Considering health equity factors is imperative when implementing a portal for mental health care to avoid heightening the digital divide for this patient population as well as to foster collaboration [33]. The development of approaches to bridge this divide for patients receiving mental health care should focus on strategies that promote these equitable health outcomes [34]. In total, 2 potential interventions to reduce the digital divide include further promoting family or caregiver collaboration and encouraging open review of clinical notes. As we defined family or caregiver support as a crucial component of the interprofessional team, a future area of exploration includes how perceptions of this support relate to use of the portal and interprofessional collaboration. Within the demographic questionnaire, the average level of family support, as rated by the participants, was 57.5 on a 0 to 100 scale. With regard to mental health care, the impact of family support has been explored in recent literature related to mental health outcomes for lesbian, gay, bisexual, transgender, queer, and similar minority youth [35] and disaster recovery for children [36]. However, this has not been extensively explored in the literature related to digital mental health interventions, such as portal use, or from the perspective of including the patient or family members as a part of the team. Furthermore, Reed et al [37] explored some factors of engagement between family members and portal use, such as reviewing laboratory results and filling prescriptions. A more in-depth analysis of how interprofessional collaboration factors align with the engagement of family members or caregivers in portal use may provide mental health clinicians with insight for enhancing interprofessional collaboration.

Being able to view different types of notes through the use of a portal was a commonly identified area by the participants to help enhance the levels of interprofessional collaboration. This process can be facilitated through the OpenNotes movement in mental health care, where patients and families or caregivers can collaboratively review their health care information with their clinicians to gain a further understanding of their care trajectory [9]. One way that the use of OpenNotes can improve interprofessional collaboration is through patient empowerment and engagement [38,39], as power can be shifted and redistributed among all interprofessional team members. This may also serve to enhance the level of trust between interprofessional team members, which was a concept mentioned various times by the participants in the interviews.

Limitations

One notable limitation of the qualitative strand of this study is that the semistructured interview guide was not produced specifically to examine interprofessional collaboration. As this secondary analysis is part of a larger study, some interview questions were tailored specifically to interprofessional collaboration, whereas other interview questions examined other factors related to the portal, such as compassion and recovery. Although responses to other questions also yielded relevant findings on this topic, future work in this space may choose to focus on additional areas of interprofessional collaboration, such as role clarification [40]. Despite being a part of the CIHC National Interprofessional Competency Framework, specific elements of role clarification between clinicians, patients, and family members or caregivers were not explored. Role clarification questions may focus on understanding individual
responsibilities within an interprofessional team and being able to understand the roles of all members [12].

A limitation of the quantitative strand was the lack of validated scales that measure interprofessional collaboration or important components of this concept. Despite being aligned with the CIHC National Interprofessional Competency Framework, self-empowerment and communication may not encompass the robust elements of what defines interprofessional collaboration of clinicians, patients, and families or caregivers. Other scales have been developed that explore elements of interprofessional collaboration but only through various clinicians (eg, physicians and nurses), rather than including patients and families or caregivers as a part of the team [41]. Finally, this study was conducted in 1 mental health hospital in Canada, and most participants were White. Therefore, our findings need to be interpreted with caution, as they have limited generalizability.

Conclusions

The integration of patient portals into mental health care has been developing over the last decade to support positive health outcomes. This secondary analysis helped us explore whether interprofessional collaboration can be supplemented through the use of a portal, specifically between clinicians, patients, and family members or caregivers. Despite nonsignificant findings from the quantitative data, narrative accounts of patients who have used a portal for their mental health care described various aspects of how it contributed to different domains of interprofessional collaboration. This provides useful information for mental health clinicians when continuing to adopt patient portals in their practice. Future work that explores these concepts related to components of health equity, such as the role of enhanced family support and collaborative note sharing, can help extend our understanding of improving portal use in mental health care in the future.

Acknowledgments

This work was supported by a Strategy for Patient-Oriented Research Grant from the Canadian Institutes for Health Research (PEH-157079) and a Sigma/Canadian Nurses Foundation Grant from the Sigma Foundation for Nursing. In-kind contributions were also provided by The Centre for Addiction and Mental Health. The funders did not have any role in content development.

Authors’ Contributions

GS conceived the study design, secured funding, and supervised all phases of the study. KD and HDS participated in qualitative data analysis and wrote the first version of the manuscript. SC and CM participated in the quantitative data analysis. BL participated in data collection. KD, HDS, BL, GS, SC, and CM revised the manuscript. All authors read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Thematic analysis: the Canadian Interprofessional Health Collaborative National Interprofessional Competency Framework. [DOCX File, 31 KB - humanfactors_v10i1e44747_app1.docx]

References


Abbreviations

CAMH: The Centre for Addiction and Mental Health
CIHC: Canadian Interprofessional Health Collaborative
HCCQ: Health Care Communication Questionnaire
MHRM: Mental Health Recovery Measure
REDCap: Research Electronic Data Capture

Edited by A Kushniruk; submitted 02.12.22; peer-reviewed by D Agarwal, S Sharma, W LaMendola, MA Islam; comments to author 08.05.23; revised version received 06.06.23; accepted 17.06.23; published 19.07.23.

Please cite as:
Durocher K, Shin HD, Lo B, Chen S, Ma C, Strudwick G
Understanding the Role of Patient Portals in Fostering Interprofessional Collaboration Within Mental Health Care Settings: Mixed Methods Study
JMIR Hum Factors 2023;10:e44747
URL: https://humanfactors.jmir.org/2023/1/e44747
doi:10.2196/44747
PMID:37467024

©Keri Durocher, Hwayeon Danielle Shin, Brian Lo, Sheng Chen, Clement Ma, Gillian Strudwick. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 19.07.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
The Polarization of Clinician and Service Staff Perspectives After the Use of Health Information Technology in Youth Mental Health Services: Implementation and Evaluation Study

Sarah McKenna¹*, BA, BSc, PhD; Sarah Piper¹*, GradDipPsych, BA (Hons); William Capon¹, BSc, MBMSc; Alison Crowley²; Lucas Lira¹, MD; Haley M LaMonica¹, BSc, MA, PhD; Min Kyung Chong¹, BSc (Hons); Elizabeth Scott¹, MBBS; Ian Hickie¹, AM, MD; Frank Iorfino¹, BSc, MCMSc, PhD

¹Faculty of Medicine and Health, Brain and Mind Centre, The University of Sydney, Camperdown, Australia
²Mind Plasticity, Sydney, Australia
*these authors contributed equally

Corresponding Author:
Sarah McKenna, BA, BSc, PhD
Faculty of Medicine and Health
Brain and Mind Centre
The University of Sydney
94 Mallett Street
Camperdown, 2050
Australia
Phone: 61 0419517723
Email: sarah.mckenna@sydney.edu.au

Abstract

Background: Highly personalized care is substantially improved by technology platforms that assess and track patient outcomes. However, evidence regarding how to successfully implement technology in real-world mental health settings is limited.

Objective: This study aimed to naturalistically monitor how a health information technology (HIT) platform was used within 2 real-world mental health service settings to gain practical insights into how HIT can be implemented and sustained to improve mental health service delivery.

Methods: An HIT (The Innowell Platform) was naturally implemented in 2 youth mental health services in Sydney, Australia. Web-based surveys (n=19) and implementation logs were used to investigate staff attitudes toward technology before and after implementation. Descriptive statistics were used to track staff attitudes over time, whereas qualitative thematic analysis was used to explore implementation log data to gain practical insights into useful implementation strategies in real-world settings.

Results: After the implementation, the staff were nearly 3 times more likely to agree that the HIT would improve care for their clients (3/12, 25% agreed before the implementation compared with 7/10, 70% after the implementation). Despite this, there was also an increase in the number of staff who disagreed that the HIT would improve care (from 1/12, 8% to 2/10, 20%). There was also decreased uncertainty (from 6/12, 50% to 3/10, 30%) about the willingness of the service to implement the technology for its intended purpose, with similar increases in the number of staff who agreed and disagreed with this statement. Staff were more likely to be uncertain about whether colleagues in my service are receptive to changes in clinical processes (not sure rose from 5/12, 42% to 7/10, 70%). They were also more likely to report that their service already provides the best mental health care (agreement rose from 7/12, 58% to 8/10, 80%). After the implementation, a greater proportion of participants reported that the HIT enabled shared or collaborative decision-making with young people (2/10, 20%, compared with 1/12, 8%), enabled clients to proactively work on their mental health care through digital technologies (3/10, 30%, compared with 2/12, 16%), and improved their response to suicidal risk (4/10, 40% compared with 3/12, 25%).

Conclusions: This study raises important questions about why clinicians, who have the same training and support in using technology, develop more polarized opinions on its usefulness after implementation. It seems that the uptake of HIT is heavily influenced by a clinician’s underlying beliefs and attitudes toward clinical practice in general as well as the role of technology, rather than their knowledge or the ease of use of the HIT in question.
Introduction

Background

The development of health information technologies (HITs) has seen recent and rapid expansion to address the well-established shortcomings within the mental health system [1-3]. In Australia and globally, widespread issues persist across the mental health system at both a structural level (ie, the arrangement and operation of services) and clinical level (ie, how care is delivered to individuals), which impact the outcomes of individuals seeking mental health care [4,5]. Issues include limited access, extensive waitlists, fragmented and disconnected services, and a lack of fundamental clinical practices that ensure that individuals receive personalized care appropriate to their level of need, such as measurement-based routine outcome monitoring and care coordination [6,7]. The COVID-19 pandemic and the resulting limitations of face-to-face care have seen a further push to implement HITs within mental health care and an increased need for literature to guide this [8,9].

More specifically, there is a call for youth mental health services to implement technologies that can facilitate more personalized care through detailed assessment and tracking of multidimensional outcomes and efficient multidisciplinary care coordination [10,11]. In Australia’s most recent study of mental health and well-being, almost half (46.6%) of female individuals aged 16 to 24 years and almost one-third (31.2%) of male individuals aged 16 to 24 years had experienced symptoms of a mental disorder in the past 12 months, which is far higher than any other age group, making youth mental health care an urgent priority [12]. A primary solution has been the funding of headspace, the National Youth Mental Health Foundation, which is mandated to establish youth-friendly, highly accessible centers that provide multidisciplinary enhanced primary care [13-15]. However, longitudinal and large cohort studies of youth accessing these services have found that only a small proportion experienced significant improvement in mental health or psychosocial functioning [16,17]. Possible explanations for this include limited resources and lack of qualified staff, particularly in rural areas, limiting the capacity of services to identify and respond to emerging mental disorders early and appropriately [4,16].

Thus, youth mental health services should be better equipped to triage care options based on levels of need (such as group therapy for clients who are at a low risk and individual therapy for clients who are at a higher risk) and to address the complexity of young people’s needs through multidisciplinary care options [4,11].

The Need for Technology-Enabled Monitoring and Care

Reviews have suggested that technology-enabled routine outcome monitoring leads to improved outcomes and reduced dropout rates from mental health care systems [18-20]. These effects are particularly strong for clients who are not on track, likely because outcome monitoring enables clinicians and clients to compare treatment progress with goals more easily and adjust therapy as needed [19]. Accordingly, the Australian Productivity Commission strongly recommended that mental health services improve their ability to provide the right health care at the right time for those with mental illness, specifically emphasizing that technology should play a larger role by improving assessment and referrals [10]. Thus, there is a strong impetus for youth mental health services to implement technology platforms that can improve the personalization of care for young people.

There are few studies and sparse literature to guide the implementation of HITs within mental health care services and to detail how they can be best used and sustained within a variety of service settings. Recent reviews of existing literature on HIT have found that user engagement is a consistent problem, varies from study to study, and is generally lower in real-world settings than in research studies [3,21,22]. For example, participant adherence to internet-based cognitive behavioral therapy can range from 6% to 100% [23]. Moreover, the implementation literature typically focuses on individual uptake, whereas there is a need to address the implementation of HIT at a service level to achieve systemic improvements in assessment, triaging, and care coordination. Some existing research suggests that the uptake of HIT by mental health professionals is commonly limited by poor digital literacy, concerns about time or financial burdens, and lack of support from service leadership [19,20,24]. However, a review of 208 articles on digital mental health interventions found only 14 articles that included a description of implementation strategies and therefore could be used to inform future HIT implementation [22].

Taken together, a stronger evidence base from real-world settings is needed to guide the successful implementation of HIT in youth mental health services.

The Development of an HIT (The Innowell Platform)

The University of Sydney’s Brain and Mind Centre (the Youth Mental Health and Technology team) has developed an HIT in partnership with young people with lived experience of mental illness, their families, clinicians, and service administrators [11]. The Brain and Mind Centre Youth Model of Care underpins this solution, arguing that multidisciplinary assessment and continuous monitoring should be used to identify the underlying trajectories of mental disorders and accurately assign the different types and levels of care according to individual needs [11]. To facilitate these clinical processes, the Innowell Platform was designed as a joint partnership between the University of Sydney, PwC (Australia), and Innowell to facilitate measurement-based mental health care [25-27] by collecting, tracking, and reporting health information back to the individual and their clinicians to inform collaborative decision-making and personalized care [28,29].

Textbox 1 provides a description of the functionalities of the technology. Notably, both the...
individual and clinician can access the individual’s health information, promoting transparency of care; this is explained in detail to the client when they are invited to use Innowell. Figure 1 provides an example of the web-based questionnaire completed by the individual, and Figure 2 shows the dashboard of results available to the clinician and client.

Innowell was co-designed and implemented in various youth mental health services through Project Synergy, as has been described in detail in previous publications [25-27]. A core feature of the implementation process was co-designing implementation strategies with services through an iterative process that allowed the research team to reflexively adapt to the individual services to address unique challenges that may be present in each setting. Previous studies have outlined the framework that was used to inform this co-design process; however, there is a need to further investigate how the co-designed strategies operated within the real-world service settings and how suitable these strategies were once implemented. Accordingly, this paper describes a preliminary observational analysis of real-world HIT implementation.

Textbox 1. Description of the functionalities of the health information technology (The Innowell Platform).

- Multidimensional assessment across a range of biopsychosocial domains (e.g., depressed mood, physical health, and sleep)
- Identification of suicidal thoughts and behaviors and subsequent notification to treating clinician and service
- Immediate dashboard of results across the range of biopsychosocial domains (as collected via the multidimensional assessment)
- Algorithms to determine the severity of needs across these biopsychosocial domains
- Data tracking and web-based progress report
- Optional support person input and health information sharing
- Health priority setting whereby people can identify 3 domains of mental health and well-being they would like to work on
- Coordination of care across multidisciplinary services
- Multiple user roles tailored to clinicians, service administrators, and individuals seeking care

Figure 1. Example of the Anxiety question set within the web-based questionnaire and example of the dashboard of results from the web-based questionnaire.

ANXIETY (1 out of 1)

These questions ask about anxiety and fear. These symptoms may include panic attacks, situational anxieties, worries, flashbacks, hypervigilance or startle. Include all of your anxiety symptoms when answering these questions. For each item, select the answer that best describes your experience **over the past week.**

Approx. 3 mins to complete

Secure and confidential

Saved automatically

Question 1
In the past week, how often have you felt anxious?
- No anxiety in the past week.
- Infrequent anxiety: Felt anxious a few times.
- Occasional anxiety: Felt anxious as much of the time as not. It was hard to relax.
- Frequent anxiety: Felt anxious most of the time. It was very difficult to relax.
- Constant anxiety: Felt anxious all of the time and never really relaxed.

Question 2
In the past week, when you have felt anxious, how intense or severe was your anxiety?
- Little or None: Anxiety was absent or barely noticeable.
- Mild: Anxiety was at a low level. It was possible to relax when I tried. Physical symptoms were only slightly uncomfortable.
- Moderate: Anxiety was distressing at times. It was hard to relax or concentrate, but I could do it if I tried. Physical symptoms were uncomfortable.
Aims
This study aimed to monitor and evaluate how the HIT (The Innowell Platform) was used naturalistically within 2 mental health services to gain practical insights into how an HIT can be best implemented and sustained to improve mental health service delivery. Furthermore, this study aimed to investigate the digital readiness of mental health service staff, the use of common clinical practices, and whether these practices can be enhanced using an HIT.

Methods

Study Design
A prospective study design was used, which included the implementation of the HIT in 2 participating sites. Data were collected via web-based surveys (at a 3-month interval over a 12-month period) and implementation logs (fortnightly) to explore clinical and service perspectives on how the HIT could be best used to facilitate improved clinical processes and outcomes within the service and to measure attitudes around the use of digital technologies in mental health care.

Implementation of the HIT
The HIT was implemented in 2 participating mental health services for 12 months (both sites chose to extend the implementation without the accompanying research measures after this period). The sites included headspace Camperdown and Mind Plasticity. Headspace Camperdown is a Commonwealth government–funded, youth-friendly, and multidisciplinary service offering early-intervention mental and physical health care and vocational support to young people aged 12 to 25 years [13]. The service has 21 staff members and is located within inner-city Sydney and provides care to approximately 1200 young people per year via psychology, psychiatry, occupational therapy, general practice, and exercise physiology. Mind Plasticity, a private, specialist practice consortia, offers multidisciplinary care to individuals of all ages who require mental health support. The service is also based in inner-city Sydney and consists of 22 staff offering psychology, psychiatry, and occupational therapy as well as education support, speech pathology, and neuropsychology services. Both sites also have a mix of contractors and employed staff.

Implementation was guided by a strategy for implementation science [26], which was developed and tested through a series of Australian government–funded research studies that implemented an HIT across a range of Australian mental health services with the aim of transforming the way mental health services deliver care to individuals [26,27,29]. Implementation phases include scoping and feasibility (assessing service resources and readiness including staffing capacity and IT requirements) and co-designing and configuring the HIT content to suit the needs of the services (eg, ensuring care options offered in the HIT reflect what the services offer, reviewing suicide notification functionality, and offering education and training on the HIT).

Implementation strategies were standardized across both settings; although once implemented, the services established their own methods of using the HIT within their service, both administratively and clinically. For example, headspace Camperdown offered the HIT’s web-based questionnaire to new clients before their first face-to-face appointment with a clinician, using this feature primarily for initial assessment,
whereas Mind Plasticity offered the HIT’s web-based questionnaire to existing clients of the practice, primarily for the purpose of routine outcome monitoring. This naturalistic approach allowed researchers to observe the impact of the HIT and collect data from service staff regarding how best to use the HIT under ecologically valid conditions that reflected a real-world service setting.

**Recruitment and Informed Consent**

All service staff, including clinicians, service managers, and service administrators, were invited to participate in this study. The participation of a broad range of service staff ensured that the feedback was collected at multiple levels for each service, including both administrative and clinical stakeholders. Eligible staff were invited to participate in web-based surveys via email from a member of the research team. If the staff indicated an interest in participating, they would receive a participant information and consent form and a survey link to provide their nonidentifiable data.

**Participant Inclusion Criteria**

Potential participants were required to meet the below inclusion criteria to participate in this study.

- Current staff (e.g., clinicians, service managers, or administrators) who work at a participating mental health service
- Aged ≥18 years
- English proficiency
- Completion of the required consent processes

**Evaluation of Clinical Opinions and the HIT**

**Web-Based Surveys**

Web-based surveys were administered to the participants (clinicians, service managers, and service administrators) using the electronic data collection software REDCap (Research Electronic Data Capture; Vanderbilt University) [30]. The surveys were based on our team’s previous research evaluating the impact of HITs on mental health services across Australia [31], with survey questions adapted and added to address the aims of this study. Specifically, data were collected about current clinical practices; if the HIT supported clinical practices; and beliefs and attitudes toward the adoption of HITs within the service, including digital readiness of staff, barriers and facilitators to adoption, and feedback on outcomes (positive or negative) that resulted from the implementation of the technology.

Participants were invited to complete a baseline survey before or during the initial phases of HIT implementation. After the completion of the baseline survey, follow-up surveys were distributed to participants at 3-month intervals to compare the effect of implementing the HIT on clinical practice over 12 months. Owing to low uptake, we were only able to report the findings from the 12-month follow-up. **Multimedia Appendix 1** provides a copy of the baseline web-based survey.

**Implementation Logs**

Implementation logs were completed monthly in REDCap by an implementation officer, who was a member of the research team and whose role included supporting the implementation of the HIT within the participating services (e.g., providing educational resources, supporting the onboarding of staff to the technology, and facilitating technical support), distributing web-based surveys to staff, and collating feedback from service staff regarding the digital health technology. The implementation logs comprised questions adapted from the Quality Implementation Framework [32] and allowed us to naturalistically evaluate the extent to which implementation processes aligned with the best practice and to document the barriers or facilitators of HIT uptake. The logs were used to document observations made by the implementation officer, over the course of a year, based on fortnightly summaries of meetings; interactions; and emails from the service staff about critical steps in implementation, such as what changes were undertaken by the service to best use the technology (e.g., service pathway changes and changes in staffing or staff roles), any technical modifications required of the HIT to improve its utility, and what aspects of the HIT and its implementation have been effective or ineffective within the service (refer to **Multimedia Appendix 2** for an example of the implementation logs). Importantly, the implementation officer aimed to embed themselves within the service where possible, primarily through the attendance of service staff meetings, to ensure that the observations from the implementation of the HIT were collected from within the service, with minimal disturbance, under real-world conditions. **Table 1** provides further details on the methods by which observations were naturalistically collected by the implementation officer to complete the implementation logs.
Table 1. Methods used to observe the implementation processes.

<table>
<thead>
<tr>
<th>Method</th>
<th>Service</th>
<th>Staff involved</th>
<th>Attendance</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case review or intake meeting</td>
<td>headspace Camperdown</td>
<td>All clinical staff, service manager, and research officer</td>
<td>Weekly</td>
<td>Meetings involved collaboratively reviewing client progress and triaging recent client intakes. The HIT&lt;sup&gt;a&lt;/sup&gt; was used to display client clinical information for team discussion.</td>
</tr>
<tr>
<td>Peer review meeting</td>
<td>Mind Plasticity</td>
<td>All clinical staff, service manager, and research officer</td>
<td>Monthly</td>
<td>Meetings involved discussion and review of client or patient progress and discussion of research projects and other collaborations when relevant (including the implementation of the HIT).</td>
</tr>
<tr>
<td>Weekly administration meeting</td>
<td>Mind Plasticity</td>
<td>Practice manager and research officer</td>
<td>Weekly</td>
<td>Meeting involved an update or discussion on the progress of the HIT implementation. This included any new developments within the service, issues or challenges, questions, or feedback from staff using the HIT.</td>
</tr>
<tr>
<td>Email correspondence and other interactions</td>
<td>headspace Camperdown and Mind Plasticity</td>
<td>All service staff</td>
<td>When required</td>
<td>All service staff were provided with the research officer’s contact details and were encouraged to contact them with any questions or feedback regarding the implementation of the HIT.</td>
</tr>
</tbody>
</table>

<sup>a</sup>HIT: health information technology.

**Data Analysis**

We used web-based surveys to collect quantitative data on staff attitudes and HIT uptake. We used descriptive statistics to compare responses before and after the implementation. Given the small sample size, it was not possible to analyze the significance of this change through quantitative methods. Qualitative data captured via implementation logs were analyzed using thematic analysis techniques and a constructivist grounded theory approach [33,34], with the aim of establishing themes regarding the use and implementation of the digital health technology within the service. An implementation officer who had been embedded in both health services established an initial list of codes based on data collected from the implementation logs. This analysis focused on identifying the barriers and facilitators of HIT uptake. Subsequently, these codes were shared and discussed with an independent researcher in a face-to-face meeting, and a list of themes was established. Subsequently, the implementation officer conducted a second round of coding to establish broader patterns of meaning within each theme. The themes were again shared with the independent researcher and refined during a face-to-face meeting. A constant comparison of similarities and differences between themes was used to identify the links between themes and to condense the overlapping themes.

Our qualitative data analysis followed the constructivist grounded theory, which assumes that all knowledge is constructed by the meanings that individuals bring to data analysis [35]. As a multidisciplinary team, our existing practical and theoretical perspectives shaped the organization of data into themes; understanding these perspectives can help explain how our sensitivities shaped our interpretation of the implementation process. The primary coder (SP) is an implementation officer who has experience working alongside youth mental health services in Australia to enhance the uptake of HITs and has a strong understanding of implementation science. The secondary coder (SM) was a clinical psychologist and academic researcher experienced in working with young people in a clinical role in youth mental health settings. Implementation science emphasizes the systemic processes that facilitate or limit the use of technology platforms in health settings. Psychological perspectives emphasize that organizational processes are underpinned by interpersonal dynamics linked to the cognitions, attitudes, and beliefs of staff within the service. Again, these perspectives informed the organization of the data into themes.

**Ethics Approval**

Ethics approval was obtained from the Executive Ethical Review Panel of the Sydney Local Health District Human Research Ethics Committee, Concord Repatriation General Hospital (2019/ETH13172). Site-specific approval was obtained for headspace Camperdown and Mind Plasticity from The University of Sydney and Sydney Local Health District, respectively.

**Results**

**Participants and Settings**

Across the 2 participating services, 43 individuals were invited to participate in this study. Of the 43 participants, 19 (44%) consented to participate in the study and completed at least 1 web-based survey. A total 63% (12/19) female and 37% (7/19) male participants, who worked across a diverse range of disciplines, were included in this study. Table 2 presents an overview of the participants’ disciplines across participating services. Owing to limited uptake from headspace Camperdown, the results were analyzed and presented using data from both services combined.
Table 2. Participants’ disciplines across participating services\(^a\).

<table>
<thead>
<tr>
<th>Role or discipline(^a,b)</th>
<th>Service</th>
<th>Mind Plasticity (n=16), n (%)</th>
<th>headspace Camperdown (n=3), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical psychologist</td>
<td></td>
<td>2 (13)</td>
<td>1 (33)</td>
</tr>
<tr>
<td>General psychologist</td>
<td></td>
<td>4 (25)</td>
<td>N/A(^c)</td>
</tr>
<tr>
<td>Provisional psychologist</td>
<td></td>
<td>1 (6)</td>
<td>N/A</td>
</tr>
<tr>
<td>Psychiatrist</td>
<td></td>
<td>3 (19)</td>
<td>N/A</td>
</tr>
<tr>
<td>Occupational therapist</td>
<td></td>
<td>2 (13)</td>
<td>N/A</td>
</tr>
<tr>
<td>Youth access clinician</td>
<td></td>
<td>N/A</td>
<td>2 (67)</td>
</tr>
<tr>
<td>Allied health assistant</td>
<td></td>
<td>1 (6)</td>
<td>N/A</td>
</tr>
<tr>
<td>General practitioner</td>
<td></td>
<td>1 (6)</td>
<td>N/A</td>
</tr>
<tr>
<td>Mental health nurse</td>
<td></td>
<td>1 (6)</td>
<td>N/A</td>
</tr>
<tr>
<td>Service administrator</td>
<td></td>
<td>3 (19)</td>
<td>N/A</td>
</tr>
</tbody>
</table>

\(^a\) Please note that 2 participants held a dual role within the service (eg, clinical psychologist and service administrator), resulting in 21 participants.

\(^b\) The mean number of years spent in each role was 7.0 (SD 9.0) years.

\(^c\) N/A: not applicable.

Staff Beliefs, Attitudes, and Uptake of the HIT

Figures 3 and 4 display staff attitudes toward the HIT, both before and after implementation. Relative to baseline, staff attitudes toward the HIT became more polarized after the implementation. After the implementation, the staff were nearly 3 times more likely to agree or strongly agree that the HIT would improve care for their clients (3/12, 25% agreed or strongly agreed before the implementation compared with 7/10, 70% after the implementation; Figure 2). Despite this, there was also an increase in the number of staff who disagreed that the HIT would improve care (from 1/12, 8% to 2/10, 20%). There was also decreased uncertainty (from 6/12, 50% to 3/10, 30% who selected not sure or neutral) about the willingness of the service to implement the technology for its intended purpose and similar rate of increase in the number of staff who agreed and disagreed with this statement.

Simultaneously, observing the implementation of new technology in their service changed the staffs’ attitudes toward their colleagues’ clinical practice. Staff were more likely to be uncertain about whether colleagues in my service are receptive to changes in clinical processes (the percentage of staff who were not sure or neutral rose from 5/12, 42% to 7/10, 70%). They were also more likely to report that their service already provides the best mental health care (agreement and strong agreement rose from 7/12, 58% to 8/10, 80%). Regarding how the platform was being used, after the implementation, a greater proportion of participants agree or strongly agree that the HIT enabled shared or collaborative decision-making with young people under their care (2/10, 20%, compared with 1/12, 8%) and enabled clients to proactively work on their mental health care through digital technologies (3/10, 30%, compared with 2/12, 16%); including apps and e-tools other than Innowell). A greater proportion of staff also agree or strongly agree that the HIT improved their assessment of and response to suicidal risk (4/10, 40% postimplementation, compared with 3/12, 25% preimplementation).

Multimedia Appendix 3, Implementation themes and associated mitigation strategies, displays the themes extracted from the implementation logs and the associated mitigation strategies adopted by the researchers and service staff.
Discussion

Principal Findings

This study assessed the perspectives of mental health service staff on an HIT platform during and after implementation and aimed to observe and evaluate the effect of the implementation process on clinical practices. Implementation log data revealed various strategies that were used by these services to support technology implementation, including education and training, on-the-ground administrative support, staggered implementation, use in team meetings, and continuous feedback to technology developers. However, despite exposure to similar implementation strategies, we found that staff attitudes toward the technology became polarized over time, both in terms of their willingness to use the platform and their belief that others in the service would be willing to adopt HIT. Thus, it appears that implementation approaches may need to be highly individualized to clinicians, and strong leadership from service
managers and funders is needed to support the successful uptake of HIT.

Given that clinicians were exposed to the same technology and implementation strategies yet had polarized reactions to the technology, the uptake of HIT in health services may ultimately be severely influenced by factors unrelated to the HIT or implementation approach. A potential explanation derived from the current literature may be that the uptake of HIT is linked to a clinician’s existing beliefs and attitudes toward clinical practice and technology, over and above their knowledge of or the ease of use of the HIT in question [19,36]. For example, previous research has found that individual processes such as internal feedback propensity, self-efficacy, and commitment to use feedback mitigate the therapist’s use of routine outcome monitoring technology [36]. In addition, common barriers to HIT implementation are that mental health professionals are often overscheduled, lack time to implement new practices, lack confidence in the confidentiality of the data, and fear that the data will not be interpreted reliably by managers or funders [37]. In summary, future research should explore the extent to which individualizing implementation strategies for health care professionals within services can improve the overall uptake of HIT.

Alternatively, service managers, policy makers, and funders need to explore how clinicians can be supported to engage in new clinical practices and make the best use of new HITs. Previous work has found that introducing new HITs or clinical practices is most likely to be sustained when a “critical mass” of staff routinely implements the new tool in their practice [38]. This allows clinicians to become more comfortable with the HIT or intervention, see it integrated into routine practice, and access peer support for the technical and emotional aspects of implementation. Accordingly, organizational support in the form of service-wide policy change, leadership from managers, and new processes to integrate HIT in clinical practice is needed so that the staff feel positively supported by the service and their colleagues to implement new HITs [38,39].

This study has important implications for policy makers, funders, and implementation science. Services may require much more significant incentives to adapt new processes and pathways that leverage the use of HIT to improve service quality. These incentives may be financial, legal, or regulatory in nature and may also arise opportunistically, for example, when mental health services were forced to adopt telehealth owing to the COVID-19 pandemic [40]. There is strong support from leading organizations such as the Australian Productivity Commission [41] and the Institute of Medicine [42] for the widespread use of HIT in services to provide person-centered and measurement-based care. This needs to be urgently backed up by key policies that provide services with the impetus for change.

Limitations
Notwithstanding these contributions, our study had some limitations. First, only 14 participants involved in clinical care completed aspects of the web-based survey regarding their attitudes toward using HIT to enhance clinical practice, which reduces the generalizability of our findings. Recruitment issues in eHealth trials have been well documented [43]. Despite the limited sample size, the in-depth evaluation of a real-world clinical service implementing a new digital technology has provided invaluable qualitative data that reflect the real-world challenges of this work. There was low readiness among staff to use HIT; thus, the small sample size in our study may reflect a general reluctance among clinicians to adopt HIT. This creates a further impetus for researchers and clinicians to continue evaluating approaches that can facilitate the implementation and use of HIT in real-world health care settings. In addition, despite identifying various processes that were used in a naturalistic mental health service to facilitate the implementation and use of HIT, our study did not evaluate the effectiveness of these processes. This was because we adopted a prospective study design that aimed to monitor how HITs were used and implemented as well as investigate digital readiness among staff. Future research is needed to evaluate the efficacy of the approaches identified in our study in increasing the implementation and use of HIT. Finally, qualitative data collection involved observations recorded by an implementation officer on implementation logs. This approach was chosen because it allowed us to naturalistically observe barriers and facilitators of HIT implementation at the service level rather than focusing on individual clinicians’ experiences. Even so, this creates a need for future research to more rigorously evaluate the underlying beliefs and attitudes that explain clinicians’ polarized experiences with HIT implementation through qualitative methods such as semistructured interviews.

Conclusions
Overall, our findings have broader implications for the future implementation of HITs in mental health services. Clinicians exposed to the same HIT, education, and support had polarized attitudes about the use of the technology, suggesting that the uptake was linked to internalized views about clinical practice and technology rather than knowledge of or the ease of use of the platform itself.

Acknowledgments
The authors disclosed receipt of the following financial support for the research, authorship, and publication of this manuscript: this study was funded by the Bupa Health Foundation. IH was supported by a National Health and Medical Research Council Research Fellowship (511921). FI was supported by the Bill and Patricia Richie Foundation. The funding sources of this study have had no input in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and the decision to submit the manuscript for publication. The authors would like to thank all staff from the participating services for their time devoted to this study.
Data Availability
The data that support the findings of this study are available from the corresponding author upon reasonable request.

Authors' Contributions
SM, SP, WC, and FI interpreted the results and wrote the manuscript. SM created the textbox, WC created the figures. All authors (SM, SP, WC, AC, LNAL, HML, MKC, ES, IH, and FI) substantively edited and revised the manuscript; approved the final version of the manuscript; and agreed to be personally accountable for the author's own contributions and ensured that questions related to the accuracy or integrity of any part of the work, even those in which the author was not personally involved, were appropriately investigated and resolved and the resolution was documented in the literature.

Conflicts of Interest
IH is the Codirector, Health and Policy at the Brain and Mind Centre (BMC) University of Sydney, Australia. The BMC operates an early-intervention youth services at Camperdown under contract to headspace. IH has previously led community-based and pharmaceutical industry–supported (Wyeth, Eli Lilly, Servier, Pfizer, and AstraZeneca) projects focused on the identification and better management of anxiety and depression. He is the Chief Scientific Advisor to, and a 3.2% equity shareholder in, Innowell Pty Ltd. Innowell was formed by the University of Sydney (45% equity) and PwC (Australia; 45% equity) to deliver the A$30 million (a currency exchange rate of A$1=US$0.64 is applicable) Australian government–funded Project Synergy (2017 to 2020) and to lead the transformation of mental health services internationally through the use of innovative technologies. ES is a Principal Research Fellow at the BMC, The University of Sydney. She is a Discipline Leader of Adult Mental Health, School of Medicine, University of Notre Dame, and a Consultant Psychiatrist. She was the Medical Director, Young Adult Mental Health Unit, St Vincent’s Hospital Darlinghurst, until January 2021. She has received honoraria for educational seminars related to the clinical management of depressive disorders supported by Servier, Janssen, and Eli Lilly pharmaceuticals. She has participated in a national advisory board for the antidepressant compound Pristiq, which is manufactured by Pfizer. She was the National Coordinator of an antidepressant trial sponsored by Servier. All other authors declare no other conflicts of interest.

Multimedia Appendix 1
Baseline web-based survey.
[DOCX File, 74 KB - humanfactors_v10i1e42993_app1.docx ]

Multimedia Appendix 2
Implementation log.
[DOCX File, 51 KB - humanfactors_v10i1e42993_app2.docx ]

Multimedia Appendix 3
Implementation themes and associated mitigation strategies.
[DOCX File, 20 KB - humanfactors_v10i1e42993_app3.docx ]

References


Abbreviations

HIT: health information technology
REDCap: Research Electronic Data Capture

Optimizing Patient-Reported Outcome Collection and Documentation in Medical Music Therapy: Process-Improvement Study

Samuel N Rodgers-Melnick1,2, MPH, MT-BC; Seneca Block1,3, MA, MT-BC; Rachael L Rivard1,4, MPH; Jeffery A Dusek1,5, PhD

1University Hospitals Connor Whole Health, Cleveland, OH, United States
2Department of Population and Quantitative Health Sciences, School of Medicine, Case Western Reserve University, Cleveland, OH, United States
3Department of Psychiatry, School of Medicine, Case Western Reserve University, Cleveland, OH, United States
4Center for Survey and Evaluation Research, HealthPartners Institute, Minneapolis, MN, United States
5Department of Family Medicine and Community Health, School of Medicine, Case Western Reserve University, Cleveland, OH, United States

Corresponding Author:
Samuel N Rodgers-Melnick, MPH, MT-BC
University Hospitals Connor Whole Health
11100 Euclid Avenue
Cleveland, OH, 44106
United States
Phone: 1 2168447727
Email: Samuel.RodgersMelnick@UHhospitals.org

Abstract

Background: To measure the effectiveness of nonpharmacologic interventions delivered during clinical care, investigators need to ensure robust and routine data collection without disrupting individualized patient care or adding unnecessary documentation burden.

Objective: A process-improvement study was undertaken to improve documentation consistency and increase the capture of patient-reported outcomes (PROs; ie, stress, pain, anxiety, and coping) within a medical music therapy (MT) team.

Methods: We used 2 Plan-Do-Study-Act (PDSA) cycles to improve documentation processes among an MT team (13.3 clinical full-time equivalent staff). Trainings focused on providing skills and resources for optimizing pre- and postsession PRO collection, specific guidelines for entering session data in the electronic health record, and opportunities for the team to provide feedback. Two comparisons of therapists’ PRO collection rates were conducted: (1) between the 6 months before PDSA Cycle 1 (T0) and PDSA Cycle 1 (T1), and (2) between T1 and PDSA Cycle 2 (T2).

Results: Music therapists’ rates of capturing any PRO within MT sessions increased significantly ($P<.001$) from T0 to T1 and from T1 to T2 for all domains, including stress (4/2758, 0.1% at T0; 1012/2786, 36.3% at T1; and 393/775, 50.7% at T2), pain (820/2758, 29.7% at T0; 1444/2786, 51.8% at T1; and 476/775, 61.4% at T2), anxiety (499/2758, 18.1% at T0; 950/2786, 34.1% at T1; and 319/775, 41.2% at T2), and coping (0/2758, 0% at T0; 571/2786, 20.5% at T1; and 319/775, 41.2% at T2). Music therapists’ feedback and findings from a retrospective analysis were used to create an improved electronic health record documentation template.

Conclusions: Rates of PRO data collection improved within the medical MT team. Although the process improvement in this study was applied to a nonpharmacologic MT intervention, the principles are applicable to numerous inpatient clinical providers. As hospitals continue to implement nonpharmacologic therapies in response to the Joint Commission’s recommendations, routine PRO collection will provide future researchers with the ability to evaluate the impact of these therapies on pain relief and opioid use.

(JMIR Hum Factors 2023;10:e46528) doi:10.2196/46528

KEYWORDS

electronic health record; integrative medicine; music therapy; pain; quality improvement
**Introduction**

To provide evidence-based, patient-centered care to hospitalized patients, health care professionals should evaluate the impact of their interventions on patient-reported outcomes (PROs) such as stress, pain, anxiety, and patients’ ability to cope with the stressors of hospitalization. PROs are vitally important at every level of health care delivery, from understanding changes within individual patients to communicating the impact of interventions to health care teams, administrators, payors, and the global research community [1].

These PROs have become increasingly important in the wake of the opioid epidemic as hospitals shift their pain management approaches from relying on opioid medication toward promoting and providing evidence-based nonpharmacologic pain treatment in accordance with Joint Commission guidelines [2-5]. In the context of inpatient integrative therapies provided for pain relief (eg, acupuncture, massage, and music therapy [MT]), PROs are an important measure of value within the health care system. PROs demonstrate whether patients’ symptoms, quality of life, and physical function are improving in response to treatment [6,7]; facilitate shared care and decision-making with the medical team [8]; and can improve patient empowerment and overall satisfaction with health care [8]. Use of PROs within interventions allows health care professionals to identify the need for modifications in the treatment plan (eg, using an active music-making MT intervention instead of a receptive MT intervention) and determine whether further action is needed to improve patients’ self-efficacy for managing their conditions [1]. Routine PRO collection is also essential within practice-based research for evaluating the effectiveness of nonpharmacologic therapies across health care systems [6,9,10].

Despite the importance of PROs in patient care and research, implementing routine PRO collection among health care professionals has been limited by factors including clinician, staff, and patient reluctance; concerns for how the data will be used; and technology challenges related to the workflow within the electronic health record (EHR) [7]. To measure the effectiveness of nonpharmacologic therapies, such as MT, investigators need to ensure robust data collection without disrupting individualized patient care or adding unnecessary documentation burden for therapists. Previous studies have described processes for improving PRO collection within nursing [11], oncology [12], outpatient integrative health and medicine [6], and a pediatric psychology service for children with sickle cell disease [13]. However, to our knowledge, no studies have described or evaluated processes for improving PRO collection among a medical MT team.

We are currently conducting a large research project entitled *Effectiveness of Medical Music therapy Practice: Integrative Research using the Electronic health record* (EMMPIRE). EMMPIRE is an observational study with three aims: (1) a retrospective study examining single-session clinical effectiveness in hematology and oncology within an academic cancer center [14] and 8 community hospitals [15]; (2) a quality improvement initiative to improve documentation consistency and increase the routine collection of PROs; and (3) a prospective study to further understand the clinical effectiveness of MT on health care use (eg, length of stay and pain medication use) and longitudinal changes in PROs.

During the EMMPIRE Aim 1 retrospective study of over 15,000 MT sessions, the investigators identified several needs for improvement within MT documentation, including (1) adding new PROs (eg, stress and coping) to measure domains for which MT was indicated; (2) increasing rates of routine PRO collection; and (3) providing structured data entry for free-text fields related to MT session characteristics. Therefore, a process-improvement study was conducted to determine if it was possible to improve documentation consistency and increase the routine collection of PROs within a medical MT team.

**Methods**

**Design and Participants**

We implemented 2 Plan-Do-Study-Act (PDSA) cycles [16] between July and December 2020 to improve assessment, evaluation, and documentation processes among an MT team. PDSA cycles are valuable quality improvement tools designed to (1) establish a plan for change, (2) execute that plan, (3) evaluate the outcome of the intervention, and (4) develop a final plan through the synthesis of the information generated during the process [17]. Our PDSA cycles (see Figure 1) addressed common barriers to routine PRO collection and documentation during MT sessions, with the primary goal to continually increase the proportion of PRO collection over 6 months. At the time of the process-improvement study, the MT team included 10.3 full-time equivalent (FTE) board-certified music therapists and 3 FTE MT interns (13.3 total FTE). Periodic retrospective EHR reviews were conducted to monitor therapists’ rates of PRO collection.
Figure 1. Plan-Do-Study-Act (PDSA) Cycles. Two PDSA cycles were implemented between July and December 2020 to improve assessment, evaluation, and documentation processes among a music therapy team. EHR: electronic health record; PRO: patient-reported outcomes; SPACE: stress, pain, anxiety, coping, education; UTA: unable to be assessed.

Setting
University Hospitals (UH) is a not-for-profit health system in Northeast Ohio serving the needs of more than 1.2 million unique patients annually. UH Connor Whole Health (UHCWH), a center for integrative health and medicine embedded within the UH health system, partners with UH physicians, providers, and institutes to meet the growing demand for the comprehensive treatment of chronic health conditions and overall well-being. UHCWH seeks to weave integrative health and medicine modalities throughout the fabric of the entire health system. Accordingly, UHCWH includes an expressive therapies program consisting of board-certified music therapists and art therapists. At the time of this study, the UHCWH Expressive Therapies Program provided MT (over 10,000 sessions per year) across 10 of UH’s 18 medical centers, including an academic medical center, a freestanding cancer center, and 8 community hospitals.

Within each of the medical centers, music therapists routinely collaborate with the medical care team (eg, physicians, advanced practice providers, nurses, social workers, and chaplains) to address patients’ symptoms and enhance psychosocial support. This program has been integrated within the clinical care team infrastructure as a nonpharmacologic resource for symptom management. Additionally, this inpatient MT program has been used to offer education (eg, verbal and written descriptions of services) on available outpatient UHCWH integrative health and medicine modalities, including chiropractic care, massage therapy, acupuncture, and integrative medicine consults.

Ethical Consideration
The EHR review procedures used in this study were approved by the UH Cleveland Medical Center institutional review board as part of a retrospective chart review with a waiver of informed consent (STUDY20191213). This study was conducted in accordance with the World Medical Association Declaration of Helsinki.

PRO Measures
The MT team was instructed to collect 0-10-point numeric rating scale (NRS) measures of stress, pain, anxiety, and coping before (presession) and after (postsession) providing an MT session. The NRS is a validated measure for acute pain intensity [18]. It has been widely used within studies of integrative therapies [10] and found to be more reliable than the visual analog scale in clinical trials, especially among patients of low socioeconomic status [10]. Investigators in previous studies have also used the 0-10 NRS to measure other domains, including anxiety in clinical effectiveness studies of nonpharmacologic interventions (eg, acupuncture, massage therapy, and meditation) [19-21] and stress in a randomized controlled trial of MT [22]. For the NRS of pain, stress, and anxiety, patients were asked, “How much (stress, pain, or anxiety) are you having right now?” with 0 signifying “none” and 10 signifying “worst possible.”

Our retrospective study revealed that coping was a common reason for MT referral [23] and a prevalent goal within MT sessions [15]. Thus, it was important to measure coping pre- and postsession to evaluate the effectiveness of MT for addressing patients’ perceived ability to cope with hospitalization. Given the challenges of implementing long, multi-item questionnaires within inpatient care [24], an NRS was chosen to conduct brief, momentary assessments of patients’ perceived coping abilities. In previous studies, investigators have used the NRS to measure changes in coping among women receiving acupuncture following mastectomy [25] and teachers [26-28]. Among teachers, the coping NRS demonstrated sensitivity to detect intervention effects [26]. In our study,
patients were asked, “How well are you coping right now?” with 0 signifying “not coping well” at all and 10 representing “coping very well.”

**Role of Research Team**

The research team leading this quality improvement initiative included a music therapist and researcher within the MT team (SRM), the manager of the Expressive Therapies Program (SB), a statistician within the UHCWH research team (RLR), and the principal investigator and director of research for UHCWH (JAD).

**PDSA Cycle 1**

**Plan**

Before this study, PRO assessment and evaluation were not established as a clinical expectation in all MT sessions. During our retrospective review of MT documentation, it was evident that several MT sessions addressing stress, pain, anxiety, and coping lacked the collection of these PROs. Furthermore, structured data entry fields within the MT EHR documentation template were used inconsistently, making it challenging to aggregate and subsequently analyze data from MT sessions (eg, format, goals, interventions, and outcomes). By providing simple tools, education, and a managerial expectation for collection of PROs, it was posited that documentation consistency and PRO collection would improve within the MT team.

**Do**

Four web-based group trainings were conducted between July and November 2020. The first training focused on setting an expectation for SPACE: collecting measures of stress, pain, anxiety, and coping and providing education about the role of MT services in the hospital. Specifically, music therapists were expected to collect either (1) pre- and postsession PROs for all MT interventions (eg, active music making, music-assisted relaxation and imagery) in which there were no patient limitations (eg, cognitive or physical limitations); or (2) presession PROs only for MT sessions in which the therapist assessed the patient and provided education but did not conduct an MT intervention. The manager of the expressive therapies program (SB) educated the MT team on techniques to approach patients and administer PRO measures verbally during a mandatory staff meeting. This education included a discussion of the importance of PROs for (1) understanding the impact of MT on individual patients; (2) communicating the impact of MT to hospital leadership; (3) investigating the real-world clinical effectiveness of MT throughout the hospital system; and (4) contributing to the evidence base for medical MT.

Additionally, the *Expressive Therapy Healing SPACE Assessment* was provided as a paper field note for therapists to use within sessions to note patient responses in real time. This field note contained (1) defined spaces for therapists to collect PRO data; (2) specific language for assessing PROs; and (3) the specific acronym expansion codes to use when documenting in the EHR. This field note was formatted to match the layout of the MT EHR documentation template so that therapists could easily transfer information from the paper field note to the EHR.

At the time of the study, radio button fields for pain and anxiety NRS were built within the MT EHR documentation template. Since changes to EHR documentation take several months, strategies were implemented to improve MT documentation using acronym expansion codes and free-text paragraph fields within the MT EHR documentation template. The acronym expansion codes (ie, expresspre, StressPre, ..., and CopingPost) created defined spaces in the narrative for therapists to enter pre- and postsession PROs. These PROs could then be mined by departmental data analysts using regular expression functions within commercial statistical packages. A step-by-step screenshot example of all procedures was included in an EHR documentation guide that was used to organize virtual team trainings. The EHR documentation guide was continually updated using therapists’ feedback and made available as a web-based resource for the MT team.

**Study**

A retrospective EHR review was undertaken to determine if the proportion of PRO collection had improved. Using clinical performance management tools within the EHR, all MT documents written during the retrospective study period and the first series of trainings were extracted. The extract provided pain and anxiety PROs within specific reportable fields. Then, regular expression functions in RStudio (version 1.3.1073) [29] were used to extract the stress and coping PROs from free-text paragraph fields. The analysis demonstrated that although the overall PRO collection rates had improved, rates of collecting stress, anxiety, and coping PROs were still less than 50% among all documented MT sessions. In addition to the PROs, there were inconsistencies in documenting (1) conflicts of service (ie, an attempt was made to see a patient but a session did not occur due to the patient being away from their room, asleep, busy, etc); (2) sessions in which the music therapist assessed the patient and provided education but did not conduct an MT intervention; and (3) sessions in which the patient fell asleep in response to an MT intervention.

In addition, 2 feedback sessions were conducted with the MT team to discuss barriers and facilitators to PRO collection and EHR documentation. Facilitators included (1) use of the field note, which provided a concrete reminder to collect PROs and a formatting structure that facilitated seamless data entry in the EHR; (2) having a laminated form patients could use to circle their PROs; and (3) discussing postsession PROs with patients to allow them to recognize their responses to MT. The MT team identified barriers to routine PRO collection among patients who were frustrated, withdrawn, or tangential in conversation. The MT team noted, importantly, that it was not possible to collect PROs among patients experiencing cognitive impairment, emotional distress, or certain physical limitations (eg, tracheostomy or sedation). The MT team expressed a desire to account for those sessions in which PROs are unable to be assessed (UTA) due to these patient limitations. Some MT team members also discussed challenges incorporating NRS measures within their therapeutic style and routine verbal skills for assessment and rapport building within MT sessions.

https://humanfactors.jmir.org/2023/1/e46528

JMIR Hum Factors 2023 | vol. 10 | e46528 | p.863

(page number not for citation purposes)
**Act**

After reviewing the MT documentation in PDSA Cycle 1 and receiving feedback from the MT team, a plan was established to capture instances in which PROs were UTA and reinforce training on documentation procedures. Additional trainings were also planned to provide instruction on how to incorporate PRO collection within routine verbal skills for assessment and rapport building, especially among patients who are frustrated, withdrawn, or tangential in conversation.

**PDSA Cycle 2**

**Plan**

Based on the knowledge gained from PDSA Cycle 1, it was posited that further improvements in documentation consistency and PRO collection were possible through (1) additional training on verbal skills for PRO collection and EHR documentation; (2) accounting for instances of outcomes UTA in rates of PRO collection; and (3) providing feedback to the MT team on how PRO collection can contribute to greater understanding and appreciation of MT’s clinical effectiveness for reducing symptom burden and improving coping.

**Do**

Four additional web-based group trainings were delivered between November and December 2020. The first training reinforced specific guidance for EHR documentation including how to document sessions in which (1) there was a conflict of service; (2) only assessment and education were provided; or (3) patients fell asleep in response to MT. Like the methods for documenting stress and coping, an acronym expansion code was created for documenting one of six reasons for outcomes UTA: (1) not applicable (the outcome was not applicable within the MT session); (2) cognitive limitation (the patient’s cognitive limitations such as dementia, confusion, or agitation prevented the patient from providing NRS); (3) physical limitation (the patient’s physical limitations such as tracheostomy, sedation, or aphasia prevented the patient from providing NRS); (4) declined (the patient declined to rate NRS); (5) emotional distress (the patient was in too much emotional distress to provide NRS); and (6) other (the MT was unable to collect NRS for some other reason such as the session being interrupted). The updated field note with UTA codes was provided to therapists to reinforce these changes.

In subsequent trainings, feedback was provided to the MT team on how the proportion of PROs collected had increased. Mean presession, postsession, and change scores were presented, demonstrating the clinically meaningful impact of MT on PROs (eg, greater than 2-unit [30,31] reductions in pain, stress, and anxiety within a single MT session). Members of the MT team shared their strategies for collecting PROs among patients who were frustrated; were withdrawn or hesitant to communicate; could not speak but could communicate in other ways; were tangential in conversation; or had challenges understanding the purpose or meaning of the PRO. Additionally, the expressive therapies program manager provided individualized feedback, as needed, to members of the MT team who had lower rates of PRO collection. This individualized feedback included a discussion of practical language skills that the therapist could integrate within their routine verbal assessment and evaluation strategies used within MT sessions.

**Study**

The processes detailed above within the study section of PDSA Cycle 1 were repeated to determine rates of PRO collection following the second series of trainings. A substantial improvement in rates of PRO collection was seen across all outcomes. During the feedback sessions, members of the MT team discussed the importance of incorporating PRO assessment within their therapeutic style and asking patients to focus on the present moment when rating their stress, pain, anxiety, and coping. One area of challenge was respondents to patients who asked why these PROs were being assessed. In these situations, the MT team recommended discussing the importance of the outcomes information for understanding how the patient was feeling and responding to the MT intervention.

**Act**

In reviewing the documentation from EMMPRIE Aims 1 and 2, it became clear that additional modifications to the MT EHR documentation template were needed. Therefore, a new MT EHR documentation template proposal was created that included the following modifications: (1) adding new radio button fields for conflict of service reason, session type (ie, 1-on-1 or group), intervention delivery (ie, in-person or digital), stress NRS, coping NRS, nausea NRS, and the Faces, Legs, Activity, Cry, Consolability scale for assessing pain behavior; (2) converting free-text fields (eg, session goal, MT interventions, and UTA) to checkbox fields for improved data clarity; (3) incorporating items from the inpatient psychiatry group flowsheet within the MT EHR documentation template; and (4) incorporating branching logic to enable the MT EHR documentation template to populate based on the type of session provided. This change also minimized the number of extra fields the therapist would need to complete in charting outcomes.

**Data Analysis**

The analysis compared therapists’ rates of capturing PROs during 3 discrete periods of MT documentation: (1) 2758 MT sessions documented in the 6 months before PDSA Cycle 1 (T0); (2) 2786 MT sessions documented during the 4 months of PDSA Cycle 1 (T1); and (3) 775 MT sessions documented during the 2 months of PDSA Cycle 2 (T2). For each period, patterns of collecting each PRO (ie, stress, pain, anxiety, and coping) were coded based on 6 different types of PRO completion, which are color-coded in Figure 2. Each session was coded as either having (1) complete pre- and postsession data; (2) presession data and the patient fell asleep in response to MT; (3) presession data only; (4) postsession data only; (5) no PRO data; or (6) any PRO data (either presession, postsession, or both). The completion proportion rates were calculated as the total number of sessions with the 6 PRO codes data divided by the total number of MT sessions. A Fisher exact test was used to compare therapists’ rates of collecting any PRO data and complete pre- and postsession data (1) between T0 and T1 and (2) between T1 and T2.
Figure 2. Patient-reported outcome completion rates among the music therapy team. T0 represents 2758 MT sessions documented in the 6 months before Plan-Do-Study-Act (PDSA) Cycle 1. T1 represents 2786 MT sessions documented during the 4 months of PDSA Cycle 1. T2 represents 775 MT sessions documented during the 2 months of PDSA Cycle 2. “*” indicates a statistically significant difference (\(P<.001\); Fisher exact test) between rates of patient-reported outcome (PRO) collection at T0 and T1. “**” indicates a statistically significant difference (\(P<.001\), Fisher exact test) between rates of PRO collection at T1 and T2. MT: music therapy; PRO: patient-reported outcomes; UTA: unable to be assessed.

To account for MT sessions in which PROs were UTA, counts and percentages of reasons outcomes were UTA were calculated during T2. Adjusted completion rates were also calculated where the total number of sessions with the 6 PRO codes was divided by the total number of MT sessions that did not have outcomes UTA. Finally, to understand sessions where only presession pain data were collected at T2, counts and percentages of sessions were tallied where (1) only MT assessment and education were provided; (2) the presession pain score was 0/10 (ie, no pain intensity); and (3) the therapist noted there was a viable reason the postsession score could not be collected.

Results

Figure 2 provides a graphical depiction of PRO completion rates among the MT team. Therapists’ rates of capturing any PRO within MT sessions increased significantly (\(P<.001\)) from T0 to T1 and from T1 to T2 for all domains, including stress (4/2758, 0.1% at T0; 730/2786, 26.2% at T1; and 298/775, 38.5% at T2), pain (482/2758, 17.5% at T0; 1022/2786, 36.7% at T1; and 344/775, 44.4% at T2), anxiety (351/2758, 12.7% at T0; 705/2786, 25.3% at T1; and 299/775, 38.6% at T2), and coping (0/2758, 0% at T0; 411/2786, 14.8% at T1; and 244/775, 31.5% at T2).

During T2, 106/775 (13.7%) MT sessions only had presession pain data. Of these 106 sessions, 79 (74.5%) were sessions where only assessment and education were provided, 10 (9.4%) were sessions where the presession pain score was 0/10, 10 (9.4%) were sessions where the therapist noted there was a reason the postsession score could not be collected (eg, interruption or decline), and 10 (9.4%) were sessions where a reason for the missing postsession score was not documented.

During T2, the MT team reported outcomes UTA in 295/775 (38.1%) MT sessions. Reasons outcomes were UTA included patients’ cognitive limitations (82/295, 27.8%), PROs not applicable to the MT session (45/295, 15.3%), patients declining to provide PROs (39/295, 13.2%), patients experiencing emotional distress (36/295, 12.2%), patients’ physical limitations (33/295, 11.2%), or other reasons not specified (60/295, 20.3%). After removing MT sessions in which outcomes were UTA from T2, therapists’ rates of collecting any PRO within MT sessions were as follows: stress (393/504, 78%), pain (476/547, 87%), anxiety (400/503, 79.5%), and coping (319/485, 65.8%).
Discussion

The purpose of this study was to determine if it was possible to improve documentation consistency and increase the capture of PROs within a medical MT team. The processes spanned every point of MT clinical care, including the training of the therapist, patient assessment and evaluation, and documentation. Through these processes, PRO collection improved throughout 10 medical centers among a team of 13.3 music therapists and MT interns. By placing all the data within the same MT EHR documentation template, the need for additional documentation burden was avoided (ie, documenting in multiple EHR sections or copying outcomes to a separate spreadsheet) as therapists implemented new skills to obtain PROs. Additionally, the documentation immediately improved without having to wait for the EHR team to build a new documentation template.

While PROs are essential elements in evaluating patients’ responses to nonpharmacologic therapies such as acupuncture, massage, and meditation [6], there are several clinical situations in which it is difficult or impossible for patients to provide PROs due to factors including physical, cognitive, or emotional limitations. Feedback from the MT team and our statistical analyses demonstrate the importance of accounting for these situations when evaluating rates of inpatient PRO collection. Additionally, it is important to understand sessions in which only the presession score is collected. Postsession scores are not appropriate in the context of assessment and education sessions where an MT intervention (eg, active music making, music-assisted relaxation, and imagery) is not delivered. Additionally, some therapists may not see the need to collect a postsession score when the patient reports presession stress, pain, or anxiety at 0/10. Since it is possible that patients could report worse scores (ie, >0) after any intervention, it is critical to ensure that therapists routinely attempt to collect postsession scores except when the patient has fallen asleep or presents with limitations as noted above.

Importantly, while rates of PRO collection continually improved over the course of the study, the rates never reached 100% of all MT sessions, even after adjusting for sessions in which outcomes were UTA. We recognize that therapists may inadvertently leave PROs out of their assessments or forget to document the reasons why outcomes were UTA. Even after accounting for UTA, the continued gap in PRO collection as demonstrated in Figure 2 (ie, stress 22%, pain 13%, anxiety 20.5%, and coping 34.2%) suggests that other factors may limit PRO collection. These factors could include (1) an education gap among new employees and interns; (2) the complex nature of individualized MT sessions among critically ill patients where specific UTA reason categories may be difficult to determine; and (3) documentation error within the EHR. Continuing education and monitoring will be needed to maintain high PRO collection rates. This continuing education will include the clinical expectation to either collect and document PROs as described in our first PDSA cycle or document why outcomes were UTA.

Our processes aligned with best practices recommended in previous studies of PRO implementation within health systems. These processes included (1) targeting multiple points in therapists’ treatment delivery and decision-making [17]; (2) integrating small systematic changes within clinical tools and resources that were already established within therapists’ routine practice [13]; (3) incorporating the MT team’s suggestions for improvement throughout the process [13]; (4) providing feedback to therapists regarding their PRO collection during monthly team meetings [11]; (5) emphasizing the value of PRO collection and the skills needed for PRO assessment and evaluation [7]; (6) building PROs within the MT EHR documentation template to facilitate data collection and provide data regarding patients’ symptoms to the medical team [6]; (7) engaged leadership [6]; (8) developing infrastructure to streamline PRO collection, data storage, extraction, and analysis [10]; (9) minimizing patient burden through brief NRS assessments [32]; and (10) continuously monitoring data completion and quality to ensure data could be stored, extracted, and used for research [8]. Additional recommendations for successful PRO collection are shown in Table 1.
The process-improvement initiative may not be feasible for other medical centers to facilitate PRO collection; (2) the clinical approaches can be applied within medical MT teams at other institutions; (3) the training procedures used in this study were applied to a nonpharmacologic MT intervention, the principles are applicable to numerous clinical therapists and providers (eg, nurses) in the inpatient setting.

This process-improvement study supports the feasibility of integrating standardized PRO collection within the clinical practice of nonpharmacologic therapies such as MT. Our training and documentation enhancements were effective at improving PRO data collection rates within a nonpharmacologic, medical MT team. For health care organizations, implementing quality improvement approaches such as these may yield similar increases in PRO collection with other clinical providers using pharmacologic or nonpharmacologic interventions. With the increased routine collection of PROs, health care organizations will be better able to (1) communicate the impact of their interventions to patients; (2) make decisions regarding which interventions to implement within inpatient care; and (3) demonstrate the value of nonbillable nonpharmacologic modalities such as MT.

With the Joint Commission inspiring health care systems to increase the availability of nonpharmacologic interventions (eg, acupuncture, massage therapy, meditation, and MT) for pain relief in hospitalized patients [3], routine collection of PROs will provide the opportunity to assess the effectiveness of these therapies at the individual and population level. Future research should seek to evaluate (1) whether these quality improvement approaches can be applied within medical MT teams at other health systems to improve PRO collection; (2) the clinical effectiveness of MT to hospital stakeholders, temporarily increasing meeting frequency to improve PRO collection is limited by a lack of data on UTA reasons at T0 and T1. It is possible that other researchers seeking to replicate our success may not have the equivalent access to EHR tools and documentation templates that were available to the MT team in our health system. This study did not consider other quality improvement approaches used in previous studies of PRO documentation enhancement, such as conducting formal interviews with music therapists to assess barriers and facilitators to PRO collection [33], recruiting clinical champions at the UH medical centers to facilitate PRO collection [7,34], developing video-based simulations modeling various techniques therapists could use to verbally collect and discuss PROs [34], or providing patients with a visual interpretation of PROs [35]. The number and frequency of meetings used to conduct this quality improvement initiative may not be feasible for other medical MT teams, and music therapists seeking to replicate these methods will need to consider their own capacity for holding these trainings. However, the trainings described in this report were conducted within the normal meeting schedule of the MT team. Given the importance of PROs for demonstrating the clinical effectiveness of MT to hospital stakeholders, temporarily increasing meeting frequency to improve PRO collection is justified.

Strengths of this study include having a baseline assessment of PRO collection, using PDSA cycle processes for quality improvement, monitoring PRO collection in real time, and minimizing documentation burden by capturing all data within the EHR. Furthermore, although the process-improvement

Table 1. Recommendations for incorporating patient-reported outcomes within a medical music therapy team.

<table>
<thead>
<tr>
<th>Category</th>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>EHR enhancement</td>
<td>• Request enhancements to EHR documentation early, as the process for implementing changes within the EHR can take several months.</td>
</tr>
<tr>
<td></td>
<td>• Supplement the request for EHR documentation changes with a request for an on-demand report of all documentation fields within each session.</td>
</tr>
<tr>
<td></td>
<td>• If fields that are relevant to clinical practice already exist within the EHR (eg, FLACC&lt;sup&gt;b&lt;/sup&gt; scale, pain NRS&lt;sup&gt;c&lt;/sup&gt;, observed emotional state, verbalized emotional state), request that these fields be added to your documentation template.</td>
</tr>
<tr>
<td>Training</td>
<td>• When a new process for documentation is initiated, ensure therapists are provided with clear and consistent training.</td>
</tr>
<tr>
<td></td>
<td>• Monitor documentation completion and consistency on a regular basis. Implement any retraining as needed.</td>
</tr>
<tr>
<td></td>
<td>• Maintain consistent and open communication with team members to ensure questions are addressed and suggestions for improvement are implemented.</td>
</tr>
<tr>
<td>Data collection</td>
<td>• Minimize documentation burden by capturing all data within one form.</td>
</tr>
<tr>
<td></td>
<td>• Use acronym expansions or dot phrases to create structured data entry within free-text fields if no structured data entry fields exist.</td>
</tr>
<tr>
<td></td>
<td>• When fields are not available for specific variables or outcomes, use regular expression functions within statistical packages to mine them from narrative portions of the documentation.</td>
</tr>
<tr>
<td></td>
<td>• In the pursuit of clean, discrete data on music therapy sessions, continue to provide space for therapists to write a narrative on the more qualitative and nuanced aspects of the music therapy session.</td>
</tr>
<tr>
<td></td>
<td>• Provide additional tools to facilitate therapists’ data collection such as a field note and a laminated form for patients to complete.</td>
</tr>
<tr>
<td></td>
<td>• Provide tools for therapists to document sessions in which PROs&lt;sup&gt;d&lt;/sup&gt; are unable to be assessed due to patient limitations. Account for these sessions when calculating PRO collection rates.</td>
</tr>
</tbody>
</table>

<sup>a</sup>EHR: electronic health record.  
<sup>b</sup>FLACC: faces, legs, activity, cry, consolability.  
<sup>c</sup>NRS: numeric rating scale.  
<sup>d</sup>PRO: patient-reported outcome.
effectiveness of medical MT using the PROs documented in the EHR; and (3) the potential for subsequent decreased opioid use and length of stay.

Acknowledgments
This work was supported by the Kulas Foundation (grant K19064R). We thank the patients analyzed in this study who received the music therapy services. We would also like to thank the following collaborators on this study: The UH Information Technology Senior Developers Carl Langdon and Ryan Jaskolka; and all participating music therapists and music therapy interns. We especially appreciate the critical support of the Kulas Foundation and its representatives: Nancy McCann, Anita Louise Steele, MM, MT-BC, and Deforia Lane, PhD, MT-BC.

Data Availability
The data sets generated and analyzed during this study are not publicly available due to privacy restrictions as the databases contain information that could compromise the privacy of research participants. However, the deidentified data sets are available from the corresponding author on reasonable request.

Conflicts of Interest
None declared.

References


Abbreviations

EHR: electronic health record
EMMPIRE: Effectiveness of Medical Music therapy Practice: Integrative Research using the Electronic health record
FTE: full-time equivalent
MT: music therapy
NRS: numeric rating scale
PDSA: Plan-Do-Study-Act
PRO: patient-reported outcome
SPACE: stress, pain, anxiety, coping, education
UH: University Hospitals
UHCWH: University Hospitals Connor Whole Health
UTA: unable to be assessed
Alerts and Collections for Automating Patients’ Sensemaking and Organizing of Their Electronic Health Record Data for Reflection, Planning, and Clinical Visits: Qualitative Research-Through-Design Study

Drashko Nakikj¹, PhD; David Kreda¹, BA; Nils Gehlenborg¹, PhD
Department of Biomedical Informatics, Harvard Medical School, Harvard University, Boston, MA, United States

Corresponding Author:
Nils Gehlenborg, PhD
Department of Biomedical Informatics, Harvard Medical School, Harvard University
10 Shattuck Street
Suite 514
Boston, MA, 02115
United States
Phone: 1 857 272 5075
Email: nils@hms.harvard.edu

Abstract

Background: Electronic health record (EHR) data from multiple providers often exhibit important but convoluted and complex patterns that patients find hard and time-consuming to identify and interpret. However, existing patient-facing applications lack the capability to incorporate automatic pattern detection robustly and toward supporting making sense of the patient’s EHR data. In addition, there is no means to organize EHR data in an efficient way that suits the patient’s needs and makes them more actionable in real-life settings. These shortcomings often result in a skewed and incomplete picture of the patient’s health status, which may lead to suboptimal decision-making and actions that put the patient at risk.

Objective: Our main goal was to investigate patients’ attitudes, needs, and use scenarios with respect to automatic support for surfacing important patterns in their EHR data and providing means for organizing them that best suit patients’ needs.

Methods: We conducted an inquisitive research-through-design study with 14 participants. Presented in the context of a cutting-edge application with strong emphasis on independent EHR data sensemaking, called Discovery, we used high-level mock-ups for the new features that were supposed to support automatic identification of important data patterns and offer recommendations—Alerts—and means for organizing the medical records based on patients’ needs, much like photos in albums—Collections. The combined audio recording transcripts and in-study notes were analyzed using the reflexive thematic analysis approach.

Results: The Alerts and Collections can be used for raising awareness, reflection, planning, and especially evidence-based patient-provider communication. Moreover, patients desired carefully designed automatic pattern detection with safe and actionable recommendations, which produced a well-tailored and scoped landscape of alerts for both potential threats and positive progress. Furthermore, patients wanted to contribute their own data (eg, progress notes) and log feelings, daily observations, and measurements to enrich the meaning and enable easier sensemaking of the alerts and collections. On the basis of the findings, we renamed Alerts to Reports for a more neutral tone and offered design implications for contextualizing the reports more deeply for increased actionability; automatically generating the collections for more expedited and exhaustive organization of the EHR data; enabling patient-generated data input in various formats to support coarser organization, richer pattern detection, and learning from experience; and using the reports and collections for efficient, reliable, and common-ground patient-provider communication.

Conclusions: Patients need to have a flexible and rich way to organize and annotate their EHR data; be introduced to insights from these data—both positive and negative; and share these artifacts with their physicians in clinical visits or via messaging for establishing shared mental models for clear goals, agreed-upon priorities, and feasible actions.

(JMIR Hum Factors 2023;10:e41552) doi:10.2196/41552
KEYWORDS
patients; electronic health records; sensemaking; pattern detection; data organization; alerts; reports; collections

Introduction

Background
During the last decade in the United States, efforts have been made to allow patients to access electronic health record (EHR) data from their providers. Although big strides have been successfully made toward setting up standards and policies to do that [1,2], less progress has been made in understanding how to aid patients in making sense of their EHR data and present them in useful and actionable ways.

Patients still predominantly access their EHR data through patient portals, usually tethered to providers’ EHR systems [3]. These tools have made some accomplishments in incorporating features that allow patients to inspect their data [4]; however, these are primarily overview and look-up features that are not designed for deeper engagement in data exploration and finding important, interesting patterns independently [5]. In addition, patient portals enable access to EHR data from a single provider, mostly with poor interoperability capabilities. This results in patients not being able to access their data from multiple providers or health care systems at the same time [6]. Such limitations usually cause patients to manually aggregate their siloed data in a time-consuming and frustrating process that is often left unfinished. Consequently, there is a widespread problem of patients’ EHR data being fragmented across multiple providers and patients having difficulties in making sense of these data in their siloed patient portals. Numerous studies point out the negative consequences for patients of these issues, emphasizing overwhelmingness with the amount of available but fragmented data [6,7]. Lack of patients’ and their physicians’ awareness of existing EHR data [6,8], underwhelming features that support patients’ independent identification of patterns in their EHR data that reflect or are tied to health problems [3,9], patients’ inability to refer back to the sensemaking activities they conducted in their portals [6], and lack of capabilities for effective evidence-based communication with their physicians [6]. These shortcomings manifest in a plethora of more concrete problems: limitations in knowing the complete picture of the patients’ medical history and ongoing health issues leading to redundant and duplicate tests, medical errors, and suboptimal decision-making and treatment; the lack of self-advocacy and poor patient-provider communication bringing difficulties in explaining the problem, setting goals, providing context and evidence, and devising care plans with referable, clear, and understandable actions; and challenges in transition of care and solving complex problems that require the engagement of multiple specialists across multiple institutions. To overcome these problems, patients have to be able to efficiently aggregate and make sense of their EHR data and reliably communicate the insights to their physicians.

Although improvements in the various patient portal usefulness and usability categories may be possible to address these goals, we decided to take a novel approach for presenting patients with their EHR data. We hope that a new model of interacting with the EHR data could provide a different perspective to the patients and, thus, open the gates for more efficient and significant improvements in supporting the sensemaking of their EHR data. To achieve this, we relied on established sensemaking theories, fundamental sensemaking activities, and principles for the collaborative determination of diagnosis and treatment. Previous research has explored how patients make sense of their personal health data for disease management, anchored in the sensemaking data frame theory [10]. This theory posits that, for an open question, individuals collect relevant data that constitute a frame [11]. Within that frame, they try to find patterns that could contribute to answering the question. During this process, the frames can be updated by adding new data, eliminating existing data, or extending to new frames. Circling back to the previously mentioned work on disease management, researchers focused on diabetes and found that frames were primarily formulated to find cause-effect relationships. These frames were grounded in contextual (eg, exercise) and clinical (eg, insulin dosage) factors for the purpose of describing different ways in which they affect or could affect the outcome measures (eg, blood glucose numbers). However, finding correlations is not the only activity that is important for sensemaking of health data. Previous research particularly has focused on the basic activities patients engage in when they are trying to make sense of their patient-generated data (PGD)—extreme values, trends, and correlations, among others [10,12]. In contrast, other work has explored patients’ sensemaking activities for their EHR data from multiple providers (eg, hospitals and clinics), such as prevalence, frequency, co-occurrence, and pre-post analysis of clinical events [9]. Furthermore, researchers have emphasized the importance of PGD in communication during clinical appointments. The PGD were perceived as facilitators to set boundaries within which parsing the space for diagnosis and treatment will take place by the physician in collaboration with the patient [13,14]. Analogous to this, we could envision EHR data being used to set similar types of boundaries. Within these boundaries, patients and their physicians can engage in new types of sensemaking activities that involve EHR data.

Motivated by this background, we can offer capabilities that allow patients to organize their EHR data in collections (ie, data frames) that can be tailored to answering their specific information needs and support sensemaking regarding their health. These collections can be manually generated by the patients and be subject to independent sensemaking activities for pattern detection. However, EHR data that come from multiple providers exhibit simple patterns, which almost anybody can find and recognize, and convoluted and complex patterns that even the greatest patient experts cannot identify and interpret [8]. Therefore, it is not always clear to patients which questions to ask (ie, what patterns to look for [15]). In addition, it can be difficult to identify such patterns completely manually [16] in a process that could be very time-consuming and requires substantial medical knowledge, analytical skills, and motivation [7,8]. Therefore, a different type of data frame can be automatically generated by the system based on patterns
in the EHR data and well-established clinical guidelines. On the basis of this argument, it appears that there is a promising approach toward supporting the sensemaking of EHR data inspired by the data frame theory. However, it has only been partially addressed by contemporary solutions and existing research.

In recent years, efforts have been made to build applications that help patients make sense of their EHR data from multiple providers. Although this idea is still in its inception, interesting new solutions have emerged, such as Apple Health Records [17], iBlueButton [18], and OneRecord [19] for mobile devices and lupHealth [20] and Discovery [21] for desktop. These solutions lay out medical records by the date they were entered, type (eg, medications and laboratory test results), or provider. On the basis of this predetermined structure, they allow users to independently explore and find patterns of interest in their EHR data, such as increasing and decreasing trends in laboratory test results or vital signs, periodicity of immunizations or medications, and co-occurrence of medical events in the same day or time interval, just to name a few. Similar to patient portals, these applications feature patient-facing alerts that are mostly focused on appointment reminders, above or below normal values in laboratory test results, or medication refill [22]. Furthermore, they are not specifically designed to support ongoing, independent sensemaking of EHR data but rather to prevent immediate issues. Despite these advancements in sensemaking support, contemporary solutions have some noticeable limitations.

Existing solutions do not allow the patient to organize their EHR data based on their information needs in a personalized way—by acute health issues or ongoing problems, for example. However, organizing personal data is a key dimension of information management and highly desired among patients [23-25]. In addition, contemporary applications offer no means of referring back to the sensemaking process at later times. This leaves patients with the tedious and frustrating burden of repeatedly collecting relevant information for frequent and related information needs, repeating the inferences over those data, or recreating mental notes. Furthermore, in cases where automatic support for surfing trends and patterns in the EHR data may exist, there needs to be a way of presenting a complete landscape of these. Moreover, patients should be able to understand these automatically generated patterns and adjust them to more actionable items for everyday life scenarios. Partial understanding and addressing of these sensemaking challenges—EHR data organization and automatic pattern detection—may often result in formulating skewed impressions of the patient’s health, thus generating misconceptions that may threaten their well-being. Previous research has provided some insights into alerting and organizing personal information, but more work is necessary for applying these concepts to EHR data.

A more comprehensive and engaging approach to pattern detection and alerting the patient to ongoing issues can be found in self-monitoring applications. Notifications for meeting goals [26] or recommendations [27] guide the user to take action for health improvement based on the patterns in the data streams coming from mobile sensors, manually entered self-assessment observations, and instrument measurements. However, such sophisticated pattern detection and recommendations can be very challenging to determine and compose owing to the sparsity and incompleteness of EHR data [28]. In addition, alerting and directing the patient based on pattern detection in EHR data poses certain risks as messages need to be framed differently than in a clinical setting. This framing must leave no room for misinterpretation by a layperson and must prevent harm at the same time [29,30]. Borrowing from existing approaches and relying on the literature, special attention should be paid to the formatting and presentation of alerts based on EHR data.

Furthermore, an opportunity to overcome the data organization shortcomings in existing applications lies in the format that is used to make the EHR data available to patients. Fast Healthcare Interoperability Resources (FHIR) is an EHR data interoperability standard that has dedicated resources for each data type that can be encountered in medical practice, such as conditions, procedures, and laboratory test results [31]. Therefore, the patient’s EHR data from multiple providers could be modeled as a set of FHIR resources (ie, records that come from multiple contributors). This setup allows us to draw analogies from existing work that focused on organizing web search results [32], relevant excerpts from web pages [33], brainstorming results from large groups [34], and pictures from the web [35]. In the spirit of the sensemaking data-framing theory, the general idea behind this work is that there exist some individual pieces of facts (ie, evidence) that need to be brought together and schematized to surface some actionable meaning. Related to this, researchers identified collections of various files (analogous to records in our case) as the most convenient and well-received way to organize data for nonexperts [36,37]. Therefore, efforts should be made to leverage this opportunity and enable personalized EHR data organization for patients.

**Objectives**

In summary, there are still open questions regarding (1) how patients would welcome and engage with artificial intelligence (AI) recommendations based on patterns in their EHR data and (2) how and why patients would want to organize their EHR data to support their sensemaking.

To address these gaps, we asked the following research questions (RQs): (1) How can we meaningfully surface automatically detected patterns for making sense of EHR data from multiple providers? In what forms and to which extent would patients want to receive such automatic support? (RQ 1); (2) How can we support the organization of EHR data to suit the patient’s needs? Why and how would patients want to organize their EHR data? (RQ 2); and (3) How can these sensemaking support improvements potentially benefit the patients? (RQ 3).

To answer these RQs, we conducted an inquisitive research-through-design study with 14 participants. Presented in the context of a cutting-edge application with a strong emphasis on independent EHR data sensemaking, called Discovery, we provided patients with high-level design mock-ups. These mock-ups demonstrated the capability for personalized transformation of EHR data into problem-based structures: (1) system-generated alerts, where the system mines
the EHR data to identify data patterns that reflect potential problems and offers recommendations, and (2) manually created collections of medical records based on health issues. By presenting the participants with such capabilities to organize and look at their EHR data based on health issues or potential health problems, we unlocked their capability for answering our RQs and conducted an inquiry into understanding how these new capabilities will affect patients’ engagement with their EHR data.

In the remainder of this paper, we present our methods and findings and provide a discussion around their interpretation and contribution. Finally, we offer some design implications and conclusions.

Methods
Overview
We conducted an inquisitive research-through-design study. This study was centered on mock-ups for 2 novel features: Alerts and Collections. The primary goal of these mock-ups was to instantiate and concretize the complex concepts that these features rely on. At its core, our approach has a research-through-design direction [38] as we believe that only after looking at these mock-ups would the study participants be able to more clearly envision our ideas and contribute to answering our RQs, which would otherwise be impossible, poorly articulated, or vague. In contrast, our work also has elements of design as inquiry. Similar to the work of Rosner [39] that introduced a new way to navigate maps, we introduced new ways for patients to navigate their medical records. Analogous to the approach by Rosner [39], we left it more open-ended in terms of the concrete needs we want to address with our design. We decided to let the study participants be inspired by the novel features and tell us more about what they believe could be achieved with those features or what improvements could be made for meeting needs that are currently not or only partially addressed by our design.

In our study, and to explore the design space more broadly by invoking feedback from a variety of potential users, we used healthy participants and participants with acute and chronic illnesses who had previously evaluated Discovery. Through Discovery, these participants were already exposed to the novel idea of making sense of medical records from multiple providers and had the necessary experience to be able to think about potential improvements in the sensemaking process.

Description of Discovery
Discovery is an open-source patient-facing sensemaking support web application for EHR data that come from multiple providers [21,40]. In this context, a provider could be any institution that provides care, such as hospitals, clinics, or private practices. Discovery works with a subset of the structured EHR data from the US core standard [41], disregarding free-text clinical notes. In its current version, it only focuses on helping patients find records relevant to their questions. However, despite providing multiple specialized views to look at the data, convenient layouts, and visualizations, this process is predominantly manual. In addition, there is no support for organizing the relevant records for a given question. For a more detailed description of Discovery, please refer to the study by Nakikj et al [9], where the authors explain the data access and its features and usability.

The key reason why we used Discovery for our study is its convenient data model based on FHIR resources that we refer to as records. At the highest level, we have the record types (eg, conditions, encounters, and immunizations). For each of these, there are record subtypes, for example, immunizations (human papillomavirus quadrivalent; influenza, seasonal, injectable, and preservative free; meningococcal conjugate vaccine; and tetanus, diphtheria, and pertussis vaccine). Each record subtype can have one or more individual records that were created at different points in time. The provenance of the record is labeled with the name of the provider (institution) that they come from. At this stage of the design, Discovery does not offer more details about which particular clinician created the record. The records, being atomic and distinguishable data structures in Discovery, will be the subject of eliciting data frames through automatic pattern detection and manual organization in our advanced sensemaking features.

We want to note that the ultimate goal of a patient-centered sensemaking support tool should always consider improving the patient-provider communication. In that spirit, we hope that the features that we are gradually trying to introduce will benefit patients to obtain a better grasp of their medical records and, in doing so, have more informed and grounded communication with their providers. However, the designs we explore in this particular study and the RQs are mostly focused on supporting patients’ sensemaking rather than patient-provider communication.

Study Design
Participants
We included 14 participants who had evaluated Discovery in a previous study. They were recruited through advertisements on Craigslist. We balanced the sample so that we had a variety of participants with respect to age, gender, and medical history. The participants had to meet the following eligibility criteria: being an adult fluent in English with a working laptop or desktop computer (with a screen size of ≥13 in), stable internet connection, normal vision or well-corrected vision with glasses or lenses, no color blindness, and medical records with one or multiple providers or institutions (hospitals or private clinical practices).

Materials
Overview
We created high-level mock-ups to roughly concretize our complex and abstract ideas for pattern detection in EHR data (Alerts) and the organization of those data (Collections). Both Alerts and Collections have the ability to transform EHR data into data frames for more efficient sensemaking—the first one being an automatic approach and the second one being manual. The data frame for an alert is centered on a meaningful pattern from the EHR data that is automatically detected by the application and is not modifiable by the patient. In that sense,
the alert already completes a fundamental sensemaking task for the patient—finding a pattern in the data. In addition, the alert has a visualization of the pattern and explanation for better understanding and conveying meaning. However, the patient still has to perform further sensemaking by considering other alerts that have also been produced by the system. In contrast, the patient manually creates a data frame by putting relevant medical records in a collection. The meaningful data pattern (or patterns) within this frame is yet to be found by the patient as the collection evolves and matures. In contrast to the alerts, the patient is able to edit the contents of the collections by adding or removing records and textual notes. Similar to the assembly of alerts, the patient can rely on the collections they have created for more comprehensive sensemaking of their health situation and medical history.

To secure an inquisitive approach, the level of detail in the mock-ups was just enough to anchor and stimulate the discussion in the desired direction and provoke brainstorming at the end of the session.

**Alerts Feature**

The **Alerts** should allow patients to have access to important and yet hidden patterns in their data that are nontrivial to detect or even unobvious to look for. Driven by well-established clinical guidelines and medical knowledge bases, the alerts should surface potential issues for patients and raise their awareness of possible upcoming problems that need to be addressed. It is important to note that the alerts in Discovery are not intended to replace physicians. Rather, they serve as an advocate for patients for matters that are otherwise unobservable for them or even for their physicians. This type of advocacy is necessary because of the fragmented data across providers and poor capabilities for comprehensive, multi-provider data pattern detection in existing patient-facing applications and EHRs.

In the mock-up in Figure 1, we present a possible iteration of the design for Discovery. Conceptually, the **SenseMaker** is the new view that unifies the functionalities of the current multiple views in Discovery and is the single place where the user goes to identify interesting patterns. Although how the **SenseMaker** looks is not relevant for this study, it was important to show that the user will be able to toggle between manual foraging for patterns and automatic pattern detection—the **Alerts** feature.

---

**Figure 1.** The high-level design mock-up for the **Alerts** feature. LAC: Los Angeles County; UCLA: University of California, Los Angeles.
EHR data for patients. The alerts are organized by record types and constructed based on triplets: a pattern identified in a single data variable (for simplicity), a visualization that presents the pattern, and the message that provides explanation and recommendations for actions. For our mock-ups, we used toy examples that were manually created and not based on any actual pattern detection or recommendation generation service, and we did not use clinical guidelines or medical knowledge bases for this purpose.

In Figure 1, the Alerts feature detects a pattern in the conditions and recommends that the patient should make a physician’s appointment soon; it then looks into the history of immunizations and realizes that 2 immunizations are overdue; and, finally, inspects the various laboratory test results and notices an upward trend in one of them (eg, cholesterol measurements), plus some values, different than what is considered normal range, that deserve further attention in 2 other laboratory test result variables (eg, blood glucose and glycated hemoglobin). After clicking on the alert with an upward trend in the laboratory test result, it is automatically shown in the Details Panel on the right. Here, there is a more detailed visualization and explanation of each of the laboratory test values individually.

Collections Feature

The Collections should allow patients to organize and annotate their medical records in a way that best suits their information needs. For example, a patient dealing with high blood pressure can create a “High Blood Pressure” collection and populate it with all their high blood pressure readings from clinical visits in the past, perhaps high BMI, or any other medical record relevant to the issue. They can add notes to the individual records for context of the measurements (eg, events from everyday life and behaviors), such as stress at work or eating large meals with alcohol. With the notes, they can also capture an insight surfaced from the collection, such as trends in the data or possible correlations among variables, for example, a relationship between high BMI and high blood pressure.

This form of data organization and insight retention capability is lacking in existing solutions. Consequently, it imposes repeated and tedious sifting through medical records in an attempt to identify the relevant ones over and over again to replicate inferences and recreate mental notes for current and frequent information needs. With the collections, in contrast, patients would be able to quickly look up and access information assembled for a particular ongoing issue and have an organized medical history, with insights, for later reflection and planning.

The key idea for the Collections feature is that, as the patient explores their data in the SenseMaker, they can add or remove records from the named collections using simple artifact marking mechanisms popular for assembling information on the web (saving a photo, ie, a pin in a Pinterest collection, or bookmarking a web page or an Instagram post) [35]. In the mock-up for the Collections view (Figure 2), there is a Collection Index and Collection Inspector. The Collection Index is a nested list of all collections that the patient created and allows for quick access to a particular collection. The Collection Inspector is the place where the patient inspects, modifies, and annotates the collection selected from the index. In the example shown in Figure 2, the user created a topic about diabetes. Within that topic, the user added 3 collections: morning spikes, unstable A1C, and medications. The first one, morning spikes, was selected for previewing or editing. The Preview only displays the records in the collection organized by record type together with the notes. In the high-level mock-up in Figure 2, we can see that the patient collected 1 condition, 3 relevant vital signs, and 2 laboratory test results. We can also see that they created 2 notes that, for example, captured measurements they made with additional context explanation. To modify the collection, the patient goes to the Edit tab. There, they can delete records or add, edit, or delete notes. In case the patient wants to add new records to the collection or simply continue exploring starting from a given collection, they can pin it to the SenseMaker and jump to that view.
Figure 2. The high-level design mock-up for the Collections feature. LAC: Los Angeles County; UCLA: University of California, Los Angeles.

Procedures

Overview

We took inspiration from the existing literature on research-through-design [38] and design as inquiry [39] to formulate our approach. We also explored ideas from design studies in the medical domain that aim to enrich the use cases, attitudes, design requirements, and functionalities of patient-facing digital tools to help us formulate the questions for the study participants [46,47]. We conducted a remote study using Zoom (Zoom Video Communications) meetings. In the 60-minute study session, the researcher shared the screen with the participant and displayed the corresponding mock-ups following a script or when the participant demanded it for reminding purposes or reference points.

We reused the demographic data to characterize the participants as digital health consumers from a previous Discovery study that they had already completed (Textbox 1). As the participants were already familiar with Discovery, the researcher only spent a little time introducing the new study and proceeded with the block dedicated to obtaining feedback on the Alerts (Textbox 2), followed by the analogous block for the Collections (Textbox 2). The study session concluded with brainstorming on new sensemaking support features (Textbox 2) inspired by the previous 2 blocks.

Textbox 1. The questions for understanding the user as a digital health consumer.

- What is your age?
- How would you describe your medical history—have you been seeing physicians a lot or not?
  - Do you have any chronic conditions—anything that makes you monitor your health more closely and have more frequent physician’s visits over longer period of time?
- How many different providers and institutions have medical information about you?
- How hard would you say it is to keep track of your medical information from those providers and institution?
  - What would be the biggest barrier for doing that?
- How comfortable are you with technology?
  - Do you currently use any devices to keep track and make sense of your health and medical information? What do you like and dislike about them?
The semistructured interview for advanced sensemaking features inspired by the mock-ups: Alerts, Collections, and brainstorming for new sensemaking support features.

<table>
<thead>
<tr>
<th>Alerts feedback (20 min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensemaking driven by an artificial intelligence (AI) agent (10 min)</td>
</tr>
<tr>
<td>- How would you feel if there was some AI agent behind the scenes looking for patterns in your data and alerting you if it finds something interesting?</td>
</tr>
<tr>
<td>- How would you see your autonomy if an AI agent drives the making sense of your data instead of you?</td>
</tr>
<tr>
<td>Feedback on the Alerts after mock-up presentation (10 min)</td>
</tr>
<tr>
<td>- What is your opinion of the Alerts feature I just presented to you?</td>
</tr>
<tr>
<td>- What are some things you liked and what is something you didn’t like?</td>
</tr>
<tr>
<td>- What are some changes or improvements you’d like to see?</td>
</tr>
<tr>
<td>- Do you have any worries about how those patterns are detected? Do you worry about their reliability or missing some of the important ones?</td>
</tr>
<tr>
<td>- When would you see the most potential for this feature to help you?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Collections feedback (25 min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data organization and reflecting on previous sensemaking (10 min)</td>
</tr>
<tr>
<td>- Currently, Discovery doesn’t support any organization of the records you identify as relevant for your questions during making sense of your data. However, we would like to support that. How would you like to be able to organize your medical data as you are finding answers to your questions?</td>
</tr>
<tr>
<td>- How would you like to be able to organize your medical data on top of what Discovery supports now for data exploration?</td>
</tr>
<tr>
<td>- What are some ways in which you would like to be able to reflect on previous explorations of your medical data?</td>
</tr>
<tr>
<td>Preparing for clinical visits (5 min)</td>
</tr>
<tr>
<td>- Discovery is also meant to help you better prepare for your upcoming clinical visits. What would help you quickly reflect and remind yourself about the key points you want to cover in the visit and the evidence in support of that?</td>
</tr>
<tr>
<td>Feedback on the Collections after mock-up presentation (10 min)</td>
</tr>
<tr>
<td>- What is your opinion of the Collections?</td>
</tr>
<tr>
<td>- What are some things you liked and what is something you didn’t like?</td>
</tr>
<tr>
<td>- What are some changes or improvements you’d like to see?</td>
</tr>
<tr>
<td>- When would you see the most potential for this feature to help you?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Brainstorming for new sensemaking features (10 min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Future improvements of Discovery (10 min)</td>
</tr>
<tr>
<td>- Now that you have a better sense of where Discovery wants to move in the future, do you have a better idea of what are some features you would like to see, but are still not there?</td>
</tr>
</tbody>
</table>

Alerts Feedback
At the beginning of this block, the participants were asked an open-ended question related to how they felt about having an AI agent going through their EHR data, finding interesting patterns, and alerting the participants about them, thus taking over the driving seat in the sensemaking process (Sensemaking driven by an AI agent in Textbox 2). This was done before showing the Alerts feature mock-up to obtain an unbiased answer. Afterward, we introduced the mock-up to be more concrete about the previous idea and inspire talking points in the semistructured feedback discussion (Feedback on the Alerts after mock-up presentation in Textbox 2).

Collections Feedback
A similar approach was taken for the Collections feature. We first started an open-ended discussion around the participants’ purpose and desire to organize and structure the relevant records they identified during the sensemaking process (Data organization and reflecting on previous sensemaking in Textbox 2). Later, this was geared toward reflection and planning, especially for clinical visits (Preparing for clinical visits in Textbox 2). After that, we were more concrete by showing the mock-up for the Collections. Participants provided feedback on what they saw and how they envisioned this feature for their personal use (Feedback on the Collections after mock-up presentation in Textbox 2).
Brainstorming for New Sensemaking Support Features

After completing the 2 blocks for the Alerts and Collections, participants were asked to provide unrestricted feedback on potential improvements and suggest new advanced sensemaking support features (Textbox 2).

Data Collection

The study session was audio recorded and transcribed using a professional service [48]. Written notes were also taken during the session and combined with the transcripts for analysis.

Data Analysis

The combined transcripts and in-study notes were thematically analyzed [49]. We took the reflexive thematic analysis approach [50]—one that allows for a more organic and flexible coding process. In this approach, there is no code book, and coding can be performed by one or more researchers, where the process is framed as a collaboration rather than reaching a consensus. The codes can evolve as the analysis progresses and are ultimately grouped into themes that convey meaning and insights regarding the subject of the investigation. In our particular case, the thematic analysis involved starting by open coding the textual data by the first author of the paper; the emerging categories were discussed and reconciled in a meeting with the second and last authors. Consequently, we surfaced the needs and boundaries of the participants for automated pattern detection and organization of their EHR data as well as the most important points in the perceptions, desired improvements, and intended use of the advanced sensemaking support features—Alerts and Collections. These categories were validated for revealing insightful themes in a group meeting with other researchers unfamiliar with Discovery and modified according to the feedback to produce the final taxonomy for the results.

Ethics Approval

We obtained approval from our Harvard Faculty of Medicine institutional review board office to conduct this study (protocol number IRB20-1757).

Informed Consent and Compensation

Each of the participants signed a consent form to take part in the study and was compensated with a US $20 Amazon gift card.

Results

Participants’ Characteristics

The participants ranged from those who considered themselves healthy (6/14, 43%), those who had episodes of acute conditions (4/14, 29%), and those who had to manage one or more chronic diseases (8/14, 57%). The age range of the participants was from 20 to 53 (mean 33.43, SD 10.39; median 30) years. We included 43% (6/14) male participants and 57% (8/14) female participants in the study. All participants (14/14, 100%) had a high school education, with most (10/14, 71%) having a college degree. Only a few participants (3/14, 21%) had professions that involved data analytics. Some participants (2/14, 14%) had very few medical records with 1 or 2 providers, whereas others (7/14, 50%) had an abundance scattered among multiple providers, from 5 or 6 up to a dozen. The rest of the participants (5/14, 36%) had records in between these 2 ranges. The participants who had rare encounters with their few providers generally found that the patient portals were useful and met their very basic needs. In contrast, those who had a lot of highly fragmented data across many providers found the experience very frustrating. Remembering how the portals worked and manually pulling data together to prepare for clinical visits or just to understand their health status was reported to be very cognitively demanding and laborious. All but 1 participant (13/14, 93%) declared being very comfortable with technology and using it on a daily basis. Most participants (11/14, 79%) had some experience tracking their health data, for which they used basic applications for running or step counts. However, none of the participants had previous experience interacting with AI agents for making sense of their medical data, nor had they ever had a chance to organize their digital medical records based on their needs within a patient portal or other application.

We classified our findings into 2 broad topics: Alerts in the process of EHR data sensemaking (RQ 1 and RQ 3) and Collections as tools for EHR data organization (RQ 2 and RQ 3). For the Alerts, we identified the following themes—implications on health attitudes, determining utilization potential, classification and appropriate dosing, keeping track of alerts, and communication with providers—and, for the Collections, we identified the following themes—use cases for the collections, generating and organizing the collections, enriching collections with PGD, and communication with providers.

In the remainder of the Results section, we will report on these themes and provide 19 quotes from 10 different participants, labeled P1 to P14.

Alerts in the Process of EHR Data Sensemaking

Overview

The vast majority of participants (12/14, 86%) were open to having the Alerts in their sensemaking features arsenal. In contrast, very few (2/14, 14%) wanted to stay away from them, stating that they were taking over the task that only a physician is equipped to perform:

This actually looks really fantastic. I feel like having these alerts, just being told what you are looking out for would be useful to me. I feel like, I don’t know, I tend to forget these things and especially looking to your...just sort of these reminders so that an immunization that might be overdue, telling me that I should make an appointment with my physician for this condition I’ve been diagnosed with. I think that these would be useful to me and I can see this sort of thing just ensuring that I am going regularly to the doctor and getting preventative care done. I think also having these, I guess, abnormalities and trends in my lab work pointed out too, that’s really useful, and I think that this is stuff that I wouldn’t necessarily know to look for on my own if I were just going through my data. The way these alerts are, it does...
The idea that this software would be what would be informing me about this rather than a professional, it seems like I’d have a hard time putting my faith in it... To me, that’s the doctor’s job, that there is all this data to filter through. It’s overwhelming, so you need a professional who understands your priorities and their priorities... My fear is that because we’re relying here on AI to find likely patterns, we’re missing the nuance. [P9]

Trust and actionability of the alerts determined how much power they would have in sensemaking, and descriptors such as topic, urgency, currency, and the sentiment of the alert determined the priority and the attitude toward it. Despite the potential usefulness of the alerts, concerns were raised about how to determine the appropriate amount and frequency. Furthermore, the history of alerts was perceived to have the potential to be transformed into a knowledge repository for dealing with health issues. Finally, the alerts were expected to support context- and evidence-based patient-provider communication and establish shared mental models between the 2. A more detailed report of these findings is presented in the following subsections.

**Implications for Health Attitudes**

Participants reported several possible implications of the Alerts feature on their health attitude. They anticipated that the alerts could raise their health awareness and stimulate their proactivity but also increase their anxiety. The participants mostly agreed that the alerts could offer a comprehensive landscape of the ongoing or upcoming issues, thus increasing their awareness of their current health status or the potential directions in which it may move. Furthermore, participants welcomed the Alerts feature as a device to stimulate their proactivity, and the alerts were perceived as a powerful nudge to prevent potential exacerbated inconveniences or significant deteriorations. However, a number of participants (4/14, 29%) were mainly concerned that this feature would keep showing numerous alerts in amounts very hard or impossible to keep up with, thus creating a possibility for anxiety.

**Determining Use Potential**

In total, 2 factors, trust and actionability, influenced how much participants would use the alerts.

Participants who were open to using the Alerts feature expressed variable levels of trust in the pattern detection algorithms and the recommendations based on them. Although some (5/14, 36%) viewed this form of AI support as an opportunity to obtain an approximate and comprehensive overview of their health status, others (2/14, 14%) felt that this approach could be very unreliable and were extremely cautious about the extent to which they would rely on it:

> I would always check it. There’s so many things that you don’t rely on just one. Myself, I have that mentality, but I’m not sure. You don’t just rely on the system entirely ever; but it’s nice to see what it catches that you weren’t able, for example, to catch, or what other things that didn’t catch that caught your attention by analyzing the data yourself, by looking at the data yourself. [P2]

Participants also wanted to know why certain patterns matter, what are the descriptions and explanations of those patterns, and what is the authority behind them—well-established clinical guidelines, hospital knowledge bases, or conclusions drawn from a large number of Discovery users. In addition, they valued the visualizations as an important complement to the verbal explanations:

> If we’re assuming that this app is only being used in the US, we can say, “According to the top hospitals, research shows that the reasons why you should get your flu shots every year is because blah, blah, blah, because your body might use immunity to the flu over time, because it takes your body X amount of weeks to actually process and absorb the vaccine,” that type of thing. [P6]

Participants put a strong emphasis on the way the recommendations were formulated—what actions they could take based on the alerts that were safe and good for them. They mostly wanted clear, unambiguous messages that moved away from a strict clinical recommendation, something that only the physician should be responsible for. However, “mild” recommendations or declarations about the status of health that orient the patient and give them a sense of direction were welcomed for the most part:

> Because there’s one thing it marks an upward trend, but there’s an upward trend that’s unsafe and an upward trend that’s not as bad, so it may have not reached an unhealthy level, but it’s enough of an upward trend that you might want to keep an eye on it... [For example] I’m noticing that your cholesterol has an upward trajectory but you’re still in the normal numbers, so this is something you should talk to a doctor about but it doesn’t seem as bad... [P3]

**Classification and Appropriate Dosing**

The participants felt overwhelmed by the potential number of alerts and desired additional capabilities of structuring and organizing the alerts by topic, urgency, currency, and sentiment. Some participants (3/14, 21%) wanted to see the alerts organized by disease or condition. They felt that this organization would make it easier for them to prioritize the numerous alerts and tackle them in a more methodological way. Several participants (4/14, 29%) noted that some alerts may need immediate attention and others could be taken care of at later times, wishing for an easy distinction between the 2:

> Maybe like you were saying if something’s urgent, like I get a lab result back and it says that I’m pre-diabetic, that would be, I think, urgent, as opposed to getting a lab back, which I have, that I had low vitamin D, so I started taking vitamin D. So that would be maybe a medium alert. [P8]

A couple of participants (2/14, 14%) pointed out that, over time, some urgent alerts could become outdated or have been already taken care of and expressed the need for keeping track of the
alerts’ currency. Several participants (3/14, 21%) shared the notion that the alerts should not exclusively stress negative trends or focus only on undesired outcomes but also point out when the patient is doing well in certain aspects of their health. According to the participants, this approach should contribute to avoiding perpetual worrying and depressive sentiments associated with the Alerts feature and also provide a sense of encouragement and accomplishment, when applicable:

> It were great to see a pop-up or email saying, “Great.” It’s like, “Your cholesterol has gone down. Your BMI, like you said, is going down. You have a lower dosage of medication. Keep it up.” [P14]

Although most of the participants (12/14, 86%) liked the idea of the alerts, they expressed worries about being overwhelmed by a potentially huge number of them. To avoid this, they wanted to be able to prioritize the alerts based on the previously covered classifications (from a previous paragraph) and determine when to look at them: pushed as detected; upon opening the application or only during exploration; or once or multiple times a day, week, or month.

**Keeping Track of the Alerts**

Almost half (6/14, 43%) of the participants perceived the Alerts feature as a log of open issues that need to be taken care of. However, they also acknowledged that taking care of such issues is not a straightforward task and might involve multiple steps. For these reasons, they wanted some capabilities to track the progress toward resolving the open issues, adding notes or tags. However, a couple of participants (2/14, 14%) raised concerns about who should be providing those tags—the patient or the physician.

**Communication With Providers**

Patients felt empowered by the Alerts through which they could potentially message their providers in the future, securing enough context and evidence with the click of a button. In addition, some expected that, as there is a notion of system-validated alert, the provider would more likely pay attention to that message and respond. In many cases, participants felt that they would need professional help to assess how urgent or important a certain alert was and further validate the recommendations it provided:

> I’d like to have to be able to contact my doctor if I see an alert, just something that it hasn’t come up before, something that I haven’t discussed with her before. So, I would like to send that information to her saying, “Discovery is picking up on this. Is it something I should worry about?” [P2]

However, some participants (3/14, 21%) raised the concern that patients might start “pinging” their providers for every single alert they encountered, overwhelming the physician to the point where they start ignoring their messages, ultimately hindering their relationship.

Interestingly, a couple of participants (2/14, 14%) laid out a flipped scenario of using the alerts for communication—instead of the patient making sense of their importance and deciding what actions to take, it should be the physician who performs that curation and contacts the patient first:

> The way I see it is that the doctor would get an alert, these trends. All this stuff is fine for the doctor to get. Doctor goes through and says, “Oh, yeah. Weight is going down. That’s okay. We expected that. Immunizations, yeah. I should have the secretary call.” Or, “Let me click and invite them to an appointment.” Something like that, but the doctor’s got to be in the loop, and I think it goes to the doctor first. [P9]

Similar to the use of the alerts for messaging with their providers, participants saw their value for in-person communication during clinical visits. They regarded the alerts as on-the-spot conversation drivers that provided satisfactory context and evidence.

**Collections as Tools for Organizing EHR Data**

**Overview**

Participants embraced the notion of grouping records relevant to a particular topic or issue under named collections and recognized how they could help in raising awareness, reflection, and health tracking. However, they wished for more flexibility in generating, enriching, and organizing the collections, as well as features that would support using them for efficient patient-provider communication:

> I feel like I would just make topics just around this chronic conditions of mine. I think I would use it to keep track of say, lab work especially over time, but also how that might line up with say, even my vital signs or...and also just put my doctor visits in there to see just so I can have a complete image of a particular condition over time in all of its different aspects. [P11]

A more detailed report of these findings is presented in the following subsections.

**Use Cases for the Collections**

Participants recognized a wide range of use cases for the Collections feature, such as quick access to relevant records, reflection on and awareness of their health, and tracking their health and journaling their diseases.

First, they suggested that they would use the collections for quickly looking up the records related to a pressing issue or other questions that frequented their minds. A mapping between frequent information needs and relevant records was one of the major benefits of the collections:

> Yes. Exactly [access relevant information fast]. So, next time she logs in, then she just have to know where to go and click and then she will see everything and then she can add some annotation? [P4]

Second, they perceived the collections as a reflection vehicle for reviewing their health status and medical history by having all sorts of issues and topics well organized and documented. This also provided a repository for raising awareness of their current health status that can be easily accessed on demand:
I think that would be useful. I think that’d be really...that’d be interesting at the very least to see how my condition has evolved and also how my thoughts and records of this condition have evolved. [P11]

According to the participants, the collections can also provide a great platform for tracking their health by inserting free-text or structured notes with concrete values. They thought that these notes could be for an individual record from the collection or the entire collection. Similarly, the participants saw the potential to journal their emotions and daily experiences with the diseases on a more elaborate level using the free-text notes and potentially capture important subtleties for understanding the effects of a treatment or disease progression or tracing back events that might have affected certain outcomes:

Just to let him [the physician] know that there’s a pattern. If there’s a certain pattern of food I’m eating and then I’m having these asthma attacks. Is it because I’m in a certain environment? It is the time of year? So, I would just track what happened, because what leads up to hours prior can determine. And it’s hard to think about it, so it’s nice to have. Think about it in long term, so it’s nice to write it down and always look at it. Because then it makes more sense. [P7]

Generating and Organizing the Collections

Participants liked the possibility of grouping and organizing their records. However, some (5/14, 36%) complained that the process was extremely manual, which made it hard to determine which records to look for and how to assess their relevance. Therefore, they wished for a level of automation in identifying the relevant records for an issue or topic. Suggestions about having prepopulated collections with records and allowing the user to modify them as necessary were offered as one of the solutions:

I think a template would be extremely useful for these things and would make me a lot more likely to use the Collections then rather than having to go and populate it myself. I trust that selecting something for my chronic condition might include categories that I forget about myself, so it would make it just a lot easier and a lot faster. [P11]

The majority of the participants (9/14, 64%) thought that having more than one level of nesting would benefit the organization of the collections. This was primarily due to the complexity of certain diseases and conditions and the need to branch them further for higher granularity to tackle problems more specifically. In addition, some participants (4/14, 29%) wished to be able to link the collections. First, they believed that this is important for the reusability of the collections—some diseases might share relevant records, and thus, separate collections related to such diseases can extend a link to an existing collection that holds those records. Furthermore, because of associations between different diseases or aspects of them, they wanted to establish relationships between different collections. Finally, they wanted to capture the evolution of diseases, for example, where one disease stemmed from another:

I think I’d be curious how much control I have as a user over setting and sort of manipulating a collection versus the Discovery system itself. For instance I can think of conditions and events I have that over time have become interrelated even though they started maybe as one off things especially when I was younger or a child but have sort of morphed as adulthood has happened into semi chronic conditions. And I sort of create a nesting effect of those and create like parent-child relationship with pins or how that grouping is happening. [P13]

Enriching Collections With PGD

Although the patients felt good about organizing records in collections, they showed reservations about the records’ comprehensiveness. They said that much of the medical events happen outside of clinical visits and deserve to be captured easily and on a regular basis:

I would just give a brief explanation. Like what happened. Like, “Oh, I had an asthma attack because I was with something that I was allergic to.” Or, “I ran too much.” I’d just give a little detailed description that the hospital wouldn’t give...Where if something happened like, “Hey, they screwed up the vital signs even though they’re on here, they weren’t accurate. Because the pulse ox strokes and they forgot to change it.” I don’t know. Just little things...Also, just to jog my memory of, “Hey, this happened when I tried to eat shrimp.” [P7]

Despite the power of free-text notes, most participants (11/14, 79%) proposed entering more structured text to capture quantifiable observations based on third-party devices. These measurements, on occasion, needed to be summarized before being entered into the application. The participants also asked for interfacing to health monitoring and tracking devices that would result in a more continuous and automatic provision of data.

Some participants (3/14, 21%) went so far as to propose a special type of record for PGD that would complement the other record types. These records would be there to store not only observations and concrete quantified values but also life events, which more often than not are causes for health to take certain directions:

Well, I think you could have a record type called “patient events.”...Yeah. So I could mark, “Oh, here’s when I got married.” Then, “Oh, look. Ever since I got married, my blood pressure has been up.” [P9]

Communication With Providers

The collections were perceived as a powerful tool in preparation for clinical visits and providing contextualized and evidence-based communication with the provider.

Having the capability to organize their data in collections before clinical visits would give participants the power to prepare the topics they wanted to cover and ask the right questions without forgetting:
I think that sounds great [collections]. I think, especially in the sense of if you have an appointment it would be nice to be able to show a doctor. I want to ask you a question about this, and I think just being...having being able to log certain things on an app is helpful. But I think I even see it just in terms of being able to ask people questions. I think that’s really helpful, because I think it could empower people too...they’re looking at their records in an easy to access way and they might be able to ask questions that they wouldn’t have thought of before. [P3]

The collections were perceived to efficiently familiarize the physician with the data related to the patient’s questions during the visit and establish a shared mental model of what the priorities for the patient are and why, as well as share the progress that has been achieved since the last visit:

I like that. I like the idea [collections]. Yeah. Because sometimes it’s like when I have a doctor’s visit, a dermatologist, for example. Like, “Oh, I use this. You suggested that I do this.” Just to have all that information of whatever prescription creams and stuff she gave me last time, and how it has worked. I don’t know. I think it would be nice to have it as a backup, especially for people that have complicated conditions, I would say. [P2]

Discussion

Novelty, Methodology and Principal Findings

To the best of our knowledge, this study represents the first attempt to understand how to apply principles of the data frame sensemaking theory to support patients’ sensemaking of their EHR data from multiple providers. We explored 2 concepts related to enabling the transformation of EHR data into "frames." The first one was about automatic extraction of meaningful patterns from the EHR data—Alerts. The second one was about manual organization of the EHR data around health issues—Collections, within which patterns could be independently observed by the patient. This study showed great interest in these novel ideas but also demonstrated that there is still a long path to carefully walk for producing designs applicable to real-life scenarios.

With our research-through-design approach, we could obtain insights about patients’ needs and preferences regarding new ways of engaging with their EHR data. The richness and reliability of the answers to our RQs related to usefulness, representations, and interactions with organized EHR data were heavily conditioned by the existence of the mock-ups. By presenting mock-ups of complex and novel features such as the Collections and Alerts, we unveiled a new frontier in patients’ conceptualization and sensemaking of their EHR data. Suddenly, patients could more concretely envision seeing their data not as a list of records ordered by time, type, or care provider but organized in a way that suits their information needs using the collections and reflects the potentially emerging health problems that deserve attention based on the system-generated alerts. This transformation from a rather crude dissection of the EHR data by type, time, and care provider to a more granular and problem-oriented view is a significant shift that became reasonably tangible and graspable to the participants in the presence of the mock-ups. Furthermore, we addressed a broader need for efficient and reliable sensemaking by introducing capabilities to transform the outlook of the EHR data into frames: manually—in collections—and system-driven—in alerts. Using this approach, we also enabled patients to retain and reuse the sensemaking work performed on their EHR data. By presenting these powerful capabilities through design mock-ups, we were able to inquire into patients’ needs, uses, and preferences related to problem-organized EHR data. We observed reports of various use cases that we could not previously envision as well as unanticipated functionalities such as strong emphasis on enriching the EHR data with PGD and bringing the physician in the loop as a supervisor and validator of their sensemaking work.

To conclude, our approach provided original contributions to the biomedical informatics and human-computer interaction fields. First, it validated the design assumption that patients want to have their data organized based on their information needs regarding current health issues and ongoing medical problems. Second, the study emphasized the importance of creating an ecosystem of EHR data and PGD that live under the same umbrella and complement each other within the confines of designated, problem-based collections. Third, the study pointed out the need for automatic support in data organization, either through automatic building of the collections or automatically detecting patterns in the data that carry some health-related meaning, good (progress) or bad (deterioration). Finally, the study indicated the importance of the physician’s role as a supervisor, validator, and editor of the sensemaking work that was performed by the patient manually or assisted by the system. We believe that these findings shed a new light on the way patients want to engage with their EHR data and open new horizons for further exploration of how to address their needs.

Interpretation of the Results and Contributions

This work provides the following contributions to the fields of human-computer interaction and biomedical informatics: (1) user needs and features for automated pattern detection in the EHR data from multiple providers—Alerts; (2) user needs and features for supporting organization and schematization of EHR data—Collections; and (3) design implications for improving the Alerts and Collections, enriching the PGD around them, and using these 2 new concepts in patient-provider communication.

Most of the participants wanted well-crafted, contextualized, and pattern-based recommendations (ie, alerts that are taxonomized and prioritized for stimulating health proactivity and securing safety and actionability). They wanted the alerts to reflect not only threats of negative outcomes but also positive developments. By introducing annotations to the alerts, the participants also saw them as a platform for tracking progress in dealing with various health issues and a knowledge base for how to face similar challenges if they arise. Furthermore, the collections were regarded as a powerful tool for awareness,
reflection, and planning. However, more structure within the collections, linking between the collections, and automatically generated collections were required. Participants stated that they would use the collection notes to log feelings, daily observations, and measurements, thus contributing to health tracking and disease journaling, but requested more variety in the formats for inputting their data. Finally, both the Alerts and Collections were perceived as a great opportunity for evidence-based communication with providers and tools for establishing a shared mental model of priorities for treatment and problem-solving.

In the attempt to propose a new way of supporting the sensemaking of medical records, we avoided the traditional user-centered design approach. This common approach typically focuses on a particular patient cohort and offers a design that meets their previously explored needs. In contrast, we decided to explore what a wider audience might expect when offered to engage with the capability to create collections of their own EHR data and see alerts based on automatically recognized data patterns. As this is a radically new approach to supporting sensemaking, we wanted to use our designs as tools for inquiring about use cases and needs rather than collecting feedback for improving a design based on previously narrowly defined and thoroughly researched user needs. That said, our designs were still well grounded in the existing literature on patients’ sensemaking of health data and our previous work. However, our designs were intentionally tailored to put emphasis on exploiting the fact that, once patients see novel features, they might start using them in unpredicted ways and for a variety of unaccounted purposes. For these reasons, we used mock-ups that were rough, provocative, and less refined but designed to let the study participants’ imagination fill in the gaps as they imagined scaffolded, personal experiences.

This approach enabled us to engage the participants in the designs by addressing broader needs that target almost anybody but then give them the opportunity to envision more specific, more personal needs and use cases. For example, by allowing data to be pulled from different providers, we address the fragmentation of the medical data for the patients—patients who start seeing a new specialist or move to a different city can benefit from that. In addition, by allowing for the organization of the data in collections, we enable more tailored engagement with the data based on issues that might benefit those dealing with multiple conditions or those who have a rich medical history. Furthermore, by allowing for alerts based on the patterns in the EHR data assembled from multiple providers, we assist in raising awareness of potential current and upcoming problems even for those who might consider themselves healthy or on top of their disease management. Although these needs are concrete, they are still broad. Given the complexity of health and medical knowledge, we wanted to dig deeper and unpack many other potential use cases and needs inspired by our new concepts of looking at EHR data. To this point, our results revealed key insights that we could not account for upfront. For example, the alerts were perceived as health status descriptors, desired to present potential deterioration but also improvements in health. Furthermore, annotations were required to keep track of issue resolution and contextualization for further use of the alerts as a knowledge base. Similarly, the collections were requested to have more internal structure and interlinking to respond to the complexity, relationships, and genesis of medical issues and diseases. In addition, various formats of PGD were needed to be more expressive in providing context for various health issues. Finally, there was a strong emphasis on the fact that the Alerts and Collections should be used in the communication with the provider, raising the perspective that these features should be designed as platforms for collaboration. We believe that the inquisitive design approach helped us uncover valuable needs and will enable us to tackle the design of the Alerts and Collections features in a more informed and traditional way further down the road—targeting particular patient groups with concrete sets of well-framed needs. For these reasons, the results of this study have the element of improving the scope of the design space and offering design directions rather than pushing a concrete design forward.

We found that most of the participants wanted some form of automatic pattern detection in their EHR data to support sensemaking and, similar to other studies, they needed well-crafted, pattern-based recommendations for establishing trust in the AI [51] and securing safety and actionability [27]. Previous work has put great emphasis on how to craft user-friendly presentations for explaining complex clinical topics [52,53] and how to deliver safe actions that patients should take based on data patterns [27,54]. Our study did not dig deep enough into these areas to provide notable findings. However, an interesting point we make is the need for a deeper context in the presentation of patterns beyond the typical reference to normal or values that are not within that range [55]. Participants wanted to have a better sense of how bad their health status actually was through the significance of the values in the pattern, the relationships between those values, and the possibilities to fluctuate in another better or worse category. In addition, and in contrast to traditional approaches that focus on notifying patients about the negative side of their health status and potential threats to their well-being [22], we found that the panel of alerts should also present the areas in which the patient is doing well.

Although there were several skeptical study participants, most (12/14, 86%) showed positive sentiments toward the alerts. However, the participants in the study were not formally familiarized with potential biases of the algorithms the Alerts feature could use in the future and might not have been aware of a variety of other limitations these algorithms can pose, such as working with sparse or incomplete data. In addition, the alerts mocked up in this study used a single variable (eg, laboratory test results); however, in reality, these will also include multiple variables (eg, medications and vital signs) and demographic information (eg, age, gender, and ethnicity). Therefore, the perceptions of the alerts may change as they become fully implemented and their limitations become more apparent. Although it appears that the alerts could be a powerful concept, we need to be aware of the potential bias in the predictions they make. The fairness of AI in health care [56] has been a popular research topic, and best efforts should be made to treat various demographics and cohorts with special attention. Our current design did not account for this as it is still in the early stages;
However, tailoring the AI algorithms in the alerts to specific patient cohorts will be seriously considered for future iterations.

Aside from being transparent and objective about the AI that the system relies on, using understandable language for the patients throughout the interface and providing pervasive assistance for learning how to use the Alerts and Collections features will be very important factors for the adoption of these features. The language should be carefully designed to fully capture the meaning behind the offered interactions with the Alerts and Collections and provide patient-friendly terminology and explanations for clinically related information. Assistance in the form of tooltips and web-based, task-oriented video tutorials should also be available to patients. Although we are considering improvements in these aspects for our current designs, they were not subject to our investigation in this study.

Participants expressed a strong interest in contributing their own data to the Alerts and Collections. Previous work by Raj et al [10] provided insights into how caregivers of patients with diabetes make sense of their clinical data (eg, insulin dosage and carbohydrate intake) enriched with context (eg, location and exercise) to determine their effects on measurable outcomes (eg, blood glucose level). However, the clinical data used for this study do not carry the full meaning of the term "clinical," which is usually associated with data that originate from a care provider institution such as a hospital or private clinic. In addition, the clinical data subject to the aforementioned study were very narrow and focused only on diabetes. In contrast, we focused our attention on clinical data in the traditional sense, the data that come from the EHR systems of the providers and are not limited to any particular disease. Therefore, our work extends the idea of enabling the sensemaking of contextualized clinical data by providing notes and annotations around EHR data.

We offer valuable insights into the value and use of PGD as enrichment to medical records, organized as alerts or collections. By adding annotations for progress toward resolving the issues represented in the alerts, the participants also saw the alerts as a repository of problem-solving knowledge that can accumulate over time. Furthermore, participants recognized an opportunity to use the notes for the individual records and the collections to log feelings, daily observations, and individual or summaries of measurements, thus contributing to health tracking and disease journaling. However, they also requested more variety in the formats for inputting their data and even suggested a separate type of record for those purposes. Reflecting on this, the idea of patients logging personal information in various formats, such as visuals, images, free text, or structured notes, for making sense of their health has been a long-standing research topic and is not novel [57-59]. However, doing so in the context of enriching the individual EHRs, patterns, and collections of them to support sensemaking around them and make those data structures more actionable in real-life scenarios is new and interesting.

On the basis of the results, the participants embraced the Alerts and Collections features as context and evidence providers as well as communication drivers that are capable of establishing shared mental models. This could probably be related to the challenges in patient-provider communication. These include difficulties in setting common ground or differences in identifying the problems and prioritizing them [60] and the need for patients to have some form of expert assistance in identifying and interpreting trends in their health data [10]. With respect to this, we should consider designing the Alerts and Collections as collaboration platforms for the patient and their physician rather than focusing exclusively on how those features can support sensemaking for the patient individually. Although the idea of a collaborative approach to treatment and diagnosis through messaging between the patient and the provider is not new [61], opening an opportunity that allows the patient to initiate communication with the click of a button in which there is a curated context and evidence already in place is relatively new to the medical domain.

However, although the Alerts and Collections features are promising tools for improving patient-provider communication, there are certain concerns related to how they can be used in the real-life workflow. First, there is the question of who creates the collections and who is able to modify them. It is conceivable that both the patient and physician can initiate a shareable collection and make edits or suggestions—the patient is the one who has much more time than the physician to dig through the data and knows their problems the best; the physician is the one with expert-level medical knowledge. Second, there is the question of who will curate the alerts. There needs to be an authority other than the AI agent—the physician, most likely—who can process the alerts and provide an interpretation of how reliable and important they are as well as what their priority is. It is conceivable that the physician can create an alert that was missed by the AI agent or override an existing one that they deem wrong, irrelevant, or inaccurate. Third, it should be noted that the designs in this study did not consider free-text clinical notes. Therefore, it remains to be further explored how these might affect patients’ organization of the EHR data and communication with their physicians. This is especially important to investigate as the lexicons that patients use typically differ from the ones in the clinical setting [62,63]. Moreover, different clinical roles—physicians (general practitioners and specialists) and nurses—may use different lexicons as well [64,65], and the notes they produce have different purposes in the overall care of the patient [66,67]. Similarly, it should be further conceptualized what role may clinical notes play in raising the alerts as our current approach only considered structured EHR data. Finally, there are concerns about the possibilities of messaging the provider frequently to the extent where the patient is ignored, which may hurt the patient-provider relationship. In summary, optimizing patient-provider communication in the presence of the Alerts and Collections features will require a very careful design in the future.

We need to reiterate one more time that the Collections and especially the Alerts should be approached and designed from the perspective that they are merely tools for supporting sensemaking and decision-making, primarily for the patient but also for the physician. As such, they should never be considered a higher authority than a human expert for taking concrete medical actions. However, they do have the capability to provide context, evidence, and reminders, all very valuable information.
that could be lacking and easily missed or overlooked by physicians. With this, the Collections and Alerts should play an important role in health awareness, proactivity, reflection, planning, and advocacy for the patients and serve as context, evidence, and insight enablers for physicians.

Finally, our study was a first step toward identifying patients’ initial reactions to the Alerts and Collections features. We were mostly focused on evoking patients’ needs, exploring use case scenarios for these features, and surfacing major preferences and concerns related to their use. At this stage, we were not interested in matching patient preferences to patient profiles, so we did not obtain extensive characteristics of the study participants. However, in pursuing refined designs set forth by the directions from this study, it will be important to thoroughly consider the patients’ cognitive, knowledge, or emotional characteristics and find out how these affect the perceptions and attitudes toward the Alerts and Collections.

**Design Implications**

**Overview**

Before we proceed with the design implications, we will list 2 important changes in the modeling and naming of the Alerts and Collections. First, we can observe that the alerts and collections are, in essence, just groupings of records (ie, data frames that support sensemaking). Both are wrapped with system-generated data or PGD with very similar purposes, which allows us to treat them the same way at the core. Second, the participants associated the alerts with negative meaning but stated that the pattern detection should present the health status of the patient with both negative and positive aspects. For these reasons, we will rename Alerts to Reports, which carries a more inclusive and neutral tone.

With these adjustments, we offer design implications for (1) contextualizing the reports more deeply for increased actionability and automatically generating the collections for more expedite and exhaustive organization of the EHR data; (2) enabling PGD input in various formats to support more granular organization, richer pattern detection, and learning from experience; and (3) using the reports and the collections for efficient, reliable, and common-ground patient-provider communication.

**Improvements for Reports and Collections**

**Interpretation of the Reports**

There should be 2 main dimensions to the report: what the health status is right now and where it can go. These assessments should be put in a broader context and explain how objectively bad or good things really are with respect to a baseline. In addition, patients should be offered a sense of how hard or easy it would be to maintain the status quo, reach a deterioration point, or improve. These additional, high-level contextualizations of the reports are important for preventing unnecessary panic and providing sometimes much needed relief, motivation, and encouragement.

**Templates for Collections**

To provide automation in determining which records to include in the collections, we propose the idea of collection templates—a mapping between different conditions and records. In the first case, the system would parse the EHR data, determine all possible collections, and prepopulate them with the relevant records. In the second case, the system would allow the user to specify a title for the collection and other metadata and, based on that input, make a best guess at what should be included in the collection. In both cases, the patient should be allowed to modify the collection template to their best interests and knowledge. To support the decision-making of what should stay in the collection or be removed, the system can assign belonging confidence measures to each of the records in the collection. These measurements can also allow the patient to manipulate the precision and recall when a collection template is created.

**PGD for Data Organization and Logging of Events**

**Granular Data Organization**

We should enable features for taxonomizing, deeper nesting of, and linking between reports and collections. For example, deeper nesting of the reports and collections could allow patients to quickly get to more specific topics. In addition, it could help in prioritizing the reports, allowing the patient to focus their attention more narrowly. Furthermore, by enabling linking, we can interrelate individual collections and reuse collections of records throughout different collections. These capabilities are particularly important for diseases that share a common genesis, similar properties, symptoms, and observations. Finally, collections can extend links to reports as a starting point for building more context and collecting additional evidence.

**Annotations for Issue Resolution Progress**

We could allow patients to tag the reports and collections with progress labels in addition to the descriptive labels reported in the results explicitly: topic, urgency, currency, and sentiment. The progress labels should come from a basic taxonomy that describes where the process is in the journey toward its resolution. This small number of labels can then be visually encoded to allow patients to quickly assess progress toward addressing their issues collectively or in isolation and make sense of their priorities.

**PGD Records**

Participants wanted more structure in the data they provided and for those contributions to be treated equally to the data that come from the EHRs. For these reasons, we can dedicate special types of PGD records to life events, manually entered observations and measurements, feelings, the ability to complete tasks in daily life, or data points from devices.

**Leveraging PGD**

This modeling of the data is expected to have secondary benefits. First, with the introduction of PGD records, we are enriching the EHR data and, therefore, enabling potentially more impactful pattern detection. Second, we allow patients to track their daily lives in a structured and searchable format, which can also provide quickly accessible and extremely valuable context and evidence in clinical visits. Finally, within a timeline-based...
historical view of reports and collections, patients can see if some issues were repeating, when, and how much. By relying on the notes and issue resolution progress annotations, patients can compile and refine strategies for how to address them in the future. In contrast, we could show the evolution of a specific collection or the variety of reports on a particular topic over time. One could imagine how different symptoms, measurements, treatments, and outcomes can vary over time and how the patient has been feeling, coping, and managing the disease in response to that, all captured in the EHR data and PGD. Consequently, this feature could be a tremendously valuable portfolio to learn from previous experiences.

**Patient-Provider Communication**

Similar to previous work [10,14,60], this study showed that patients perceive the provider as a partner in their sensemaking. Future designs of patient-facing sensemaking tools should account for this partnership and provide features that enable the establishment of shared mental models and artifact-based communication.

**Context and Evidence-Based Communication**

For example, one can imagine how a report or collection can be attached to a message and have that message directly reference particular records, notes, or annotations from them. This will provide more granular contextualization for different points in the message and yet keep the full context in the report or collection if needed.

**Establishing Shared Mental Models**

In addition, the tagging of the reports and collections for detailed description and issue resolution progress could be enabled for the patient’s physician as well. With this 2-sided labeling mechanism in place, we could encourage the detection of potential discrepancies in the perceptions of whether certain issues have been addressed or are still open and what the progress is. These could then be further transformed into high-priority talking points via messaging or in clinical visits. Analogously, the collections can be collaboratively edited as well. The physician could also have the right to initiate and populate a collection or suggest adding or removing records for an existing collection, whether created by the system (collection template) or the patient.

**Message and Task Distribution Among Care Team Members**

To avoid physician burnout as a consequence of overwhelming messages, a triaging method should be put in place. For example, clear guidelines should be provided to the patient regarding which care team member should be targeted based on the content of their message. In addition, and because even well-defined guidelines can be difficult to follow, a designated care team member (other than the physician) can perform the triaging manually. Finally, each care team member should have different editing and validation privileges for the patients’ collections and reports. To set common grounds, all messaging, editing, and validating activities should be made available to the entire care team. For transparency of care, this history of activities should also be visible to the patient. By all means, special attention in the design should be paid to whether full transparency applies for all types of activities or whether some should be best left undisclosed to avoid unnecessary overhead in communication, confusion, misguidance, or worry.

**Limitations**

This study has several limitations. First, the participant sample did not include older adults, who might have different needs, perceptions, and preferences. Second, we might have introduced certain biases in the mock-ups, although the timing of their introduction and presentation was carefully tailored to avoid this. Third, a lot of the opinions of the participants were based on projecting their expectations for something they had not yet experienced in real life, such as making sense of their own EHR data from multiple providers in a single application, building a collection of records and using it in clinical visits, or making decisions based on a panel of alerts. Fourth, the insights from the participants were obtained based on mock-ups rather than on a fully functional system based on their own EHR and personally generated data, which may have skewed their perceptions or depleted the richness of their feedback. Finally, we did not obtain detailed patient characteristics to map patient profiles to specific needs and preferences.

However, we believe that we came fairly close to our goal of painting a broad picture of what patients’ needs are and how we can design features that support automatic pattern detection and EHR data organization for improved sensemaking and use in real-life scenarios.

**Conclusions**

There are untapped opportunities to support automatic pattern detection and organization of EHR data from multiple providers for patient-facing sensemaking applications. In this paper, we investigated the needs patients have with respect to these capabilities and the features that they would prefer for addressing those needs. We learned that patients are open to carefully designed automatic pattern detection with safe and actionable recommendations, which produces a well-tailored and scoped landscape of reports for both potential threats and positive progress. We also learned that patients are willing to contribute their own data in the form of notes, tags, and structured formats to enrich the meaning and enable easier sensemaking of their EHR data through reports and collections. Finally, the study showed that patients wanted to use these artifacts for raising awareness, reflection, and planning but, above all, for contextualized and evidence-based patient-provider communication via messaging or in clinical visits. These findings resulted in design implications for contextualizing more deeply the reports for increased actionability and automatically generating the collections for more expedited and exhaustive organization of the EHR data; enabling PGD input in various formats to support more granular organization, richer pattern detection, and learning from experience; and using the reports and collections for efficient, reliable, and common-ground patient-provider communication.

Although our study was nested in Discovery, the results and design implications can be easily generalized to other existing and future systems. The most important takeaway from this study is that patients need to have a flexible and rich way to organize and annotate their EHR data; be introduced to insights...
from their data—both positive and negative; and share these artifacts with their physicians during clinical visits or via messaging for establishing shared mental models for goals, priorities, and actions.

Although at this point, we have a better grasp of the direction for supporting automated sensemaking and organization of EHR data for patients, we must not forget that it will take a significant effort until we have a fully functional system. We believe that collaborative efforts and strategizing will benefit the implementation of the insights from our study. With respect to the Reports feature, it will take engagement from a wider community to assess the quality requirements for the EHR data for various individual reports, the feasibility and fairness of the data pattern detection algorithms, and the meaningfulness and understandability of the recommendations based on these patterns. Multiple research groups can try to implement different reports (eg, cardiovascular, respiratory, and renal) that can arise from widely adopted clinical guidelines or other trusted sources of knowledge. Furthermore, a gradual approach that primarily targets prevalent conditions and feasible reports should be the starting point so that we can engage wider audiences in realistic and robust evaluations to produce a broader impact. Similarly, any automatic collections should take a similar approach to the reports. For example, a research group can focus on determining which records should be included in a collection for high blood pressure, for kidney failure, and so on. These collaborative efforts can produce a pool of reports and collections that different research groups can borrow from in the implementation of their sensemaking support tools. This will potentially enable faster design-implementation-evaluation cycles and, consequently, the advancement in our knowledge about patients’ use of the reports and collections individually or in collaboration with their physicians.

Although research groups can exert tremendous efforts following the previous guidelines, a key to the success of patient-facing sensemaking tools is the involvement of clinical professionals. In fact, our study pointed out that these tools should, in principle, be regarded as collaboration platforms that improve communication and promote partnership between the patient and physician. With that said, research groups should nurture great relationships with physicians who can contribute valuable insights for the design of such tools. Moreover, to produce usable designs, we will need to engage in evaluations that take place in a clinical setting. For this, physician collaborators will be essential in securing a welcoming setup in their office and even engage as assistants in the research. For successful evaluations, physicians should be willing to sacrifice the comfort of their well-established workflows and have those interrupted by the new sensemaking tools that will inevitably be used at the point of care.

This work is focused on empowering patients by supporting their capabilities to make sense of their EHR data and make these more actionable in real-life scenarios. However, the principles from the Reports could be translated to position the health care team member as the central figure of sensemaking. With this approach, clinicians would be able to obtain reports from an AI agent on the patient’s health status that include their clinical data, demographics, and social determinants of health. In this report, the most likely options for interventions would be suggested based on patient-reported outcomes and clinical outcomes combined with established clinical guidelines. Although this approach does not attempt to shortcut the clinicians as decision makers, it should provide a variety of options to consider as brainstorming for alternatives for individuals in isolation can typically be a cognitively demanding task and result in omitting viable solution paths. In addition, this approach is well aligned with the concept of a learning health care system, which is the guiding star of the latest research endeavors. With recent advancements in AI, we should start preparing for a setting in which there are AI agents that support the work of patients and physicians. In this futuristic setup, which might not be far from now, we can expect that AI agents will help patients in their self-advocacy by assisting them in the sensemaking of their health data and communicating with their providers. In contrast, AI agents will help clinicians decide what is the best care path for the patient and how to communicate that back to them. Having a collaboration between a patient, AI agents, and clinicians will bring an interesting dynamic in the patient-provider communication, which will deserve a deep engagement from researchers. Questions of the type of when and for what tasks AI agents can improve communication, shared decision-making, and the patient-provider relationship will be of high priority.

Finally, although this study produced exciting new design directions for supporting patients’ sensemaking of their EHR data, we have to point out that the features it promotes are disruptive in nature. First, they challenge patients to change the way they interact with their EHR data. Second, they also require adjustments to the existing workflows in the clinical visit and shifting the novice-expert relationship between the patient and the physician toward partnership. We would like to stress that these changes might face amplified resistance and an extended time to take place if we are not extremely careful with our designs and respectful of existing practices. A gradual approach that involves the patients and physicians at every step of the design iteration should be taken. Carefully listening to both stakeholders to deeply understand their needs should help in finding design compromises that will benefit both parties as a team and, ultimately, contribute to a better patient experience and clinical outcomes.

Acknowledgments

This work was funded by Scripps Research through the Technology to Empower Changes in Health (TECH) Network Participant Center grant award (5U24OD023176).
Conflicts of Interest

NG is a cofounder and equity owner of Datavisyn. All other authors declare no other conflicts of interest.

References

1. New applicability dates included in ONC interim final rule. The Office of the National Coordinator for Health Information Technology. URL: https://www.healthit.gov/sites/default/files/2022-08/Highlighted_Regulatory_Dates_All.pdf [accessed 2021-08-13]

2. ONC’s cures act final rule. HealthIT.gov. URL: https://tinyurl.com/yevuzdkt [accessed 2021-08-13]


18. iBlueButton homepage. iBlueButton. URL: https://www.ibluebutton.com/ [accessed 2021-09-08]

19. Digital health platform combining your medical records in one place. OneRecord. URL: https://onerecord.com [accessed 2021-09-08]

20. 1upHealth homepage. 1upHealth. URL: https://1up.health/product/patient [accessed 2021-09-08]
40. sync-for-science / s4s-discovery. GitHub. URL: https://github.com/sync-for-science/s4s-discovery [accessed 2021-08-22]


Abbreviations
AI: artificial intelligence
EHR: electronic health record
FHIR: Fast Healthcare Interoperability Resources
PGD: patient-generated data
RQ: research question

©Drashko Nakikj, David Kreda, Nils Gehlenborg. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 21.08.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
A Web-Based Self-management App for Living Well With Dementia: User-Centered Development Study

Abigail Rebecca Lee, BSc, MSc; Emese Csipke, BA, MSc, PhD; Lauren Yates, BSc, PhD, DClinPsy; Esme Moniz-Cook, MSc, PhD, DClinPsy; Orii McDermott, ARCM, BA, MMT, PhD; Steven Taylor, BA; Michael Stephens, BSc; Daniel Kelleher, MA; Martin Orrell, BMedSci, MBBS, PhD

1Institute of Mental Health, School of Medicine, University of Nottingham, Nottingham, United Kingdom
2Division of Psychiatry, University College London, London, United Kingdom
3Department of Clinical, Educational and Health Psychology, University College London, London, United Kingdom
4Faculty of Health Sciences, University of Hull, Kingston-upon-Hull, United Kingdom
5Ayup Digital, Leeds, United Kingdom
6Research & Development, Humber Teaching National Health Service Foundation Trust, Witham, United Kingdom
7Centre for Applied Dementia Studies, University of Bradford, Bradford, United Kingdom

Corresponding Author:
Abigail Rebecca Lee, BSc, MSc
Institute of Mental Health
School of Medicine
University of Nottingham
Jubilee Campus
Triumph Road
Nottingham, NG7 2TU
United Kingdom
Phone: 44 07563254838
Email: abigail.lee19@nhs.net

Abstract

Background: Self-management, autonomy, and quality of life are key constructs in enabling people to live well with dementia. This population often becomes isolated following diagnosis, but it is important for them to feel encouraged to maintain their daily activities and stay socially active. Promoting Independence in Dementia (PRIDE) fosters social inclusion and greater dementia self-management through an interactive handbook.

Objective: This study aimed to develop a paper-based PRIDE manual on a web-based platform.

Methods: Two overarching stages were used to create the web-based version of PRIDE. The first was Preliminary Development, which encompassed tendering, preliminary development work, consultations, beta version of the website, user testing and consultation on beta version, and production of the final web-based prototype. The second stage was Development of the Final PRIDE App, which included 2 sprints and further user testing.

Results: Through a lengthy development process, modifications were made to app areas such as the log-in process, content layout, and aesthetic appearance. Feedback from the target population was incorporated into the process to achieve a dementia-friendly product. The finished PRIDE app has defined areas for reading dementia-related topics, creating activity plans, and logging these completed activities.

Conclusions: The PRIDE app has evolved from its initial prototype into a more dementia-friendly and usable program that is suitable for further testing. The finished version will be tested in a reach, effectiveness, adoption, implementation, and maintenance study, with its potential reach, effectiveness, and adoption explored. Feedback gathered during the reach, effectiveness, adoption, implementation, and maintenance study will lead to any further developments in the app to increase its applicability to the target audience and usability.

(JMIR Hum Factors 2023;10:e40785) doi:10.2196/40785
Introduction

Background
Living well with dementia has often been constructed around quality of life, choice, autonomy, dignity, and staying as independent as possible [1]. People with dementia have identified how they quantify living well, which included involvement at home and in the neighborhood, independence, self-management of symptoms, and quality of life. They also recommend that these should be considered when developing dementia-specific interventions [2]. Many people with dementia have the ability to maintain an active and social life, but some of the negative effects of receiving a diagnosis, depression, or diagnosis stigma can result in social isolation and withdrawal from society [3,4]. It is important that people living with mild dementia are supported and encouraged to maintain their normal activities, remain independent, and stay active within society for as long as they are able to.

Promoting Independence in Dementia (PRIDE) is a psychosocial program designed for people living with mild dementia, whose symptoms of dementia affect day-to-day activities, but are able to live relatively independently and to promote choice, autonomy, and social inclusion. It encourages them to maintain and develop cognitive, physical, and social activities to improve their self-management, independence, and quality of life. The content is delivered in a manualized format, with interactive activities and discussion points, such as creating activity plans. Users are paired with trained facilitators who go through the PRIDE program and support the development and execution of personalized activity plans. Across 3 sessions, users and facilitators plan, carry out, and review users’ individual plans and discuss how techniques learned through PRIDE could support them in approaching activities in the future.

A multicenter feasibility study of the PRIDE program provided participants with both a paper manual and an electronic version, so they were able to choose whether to use one or both formats [5]. The paper manual was the most popular, being used by all participants in the intervention arm, but 1 participant chose to use both the paper and the electronic versions. The findings suggested that the PRIDE intervention was a useful and relevant program to promote independence and support people with dementia in their daily activities, and it was generally well-received by the participants [5]. Although only 1 participant accessed the electronic version of PRIDE, the COVID-19 pandemic meant that more people have resorted to web-based resources; therefore, further developments to refine the PRIDE web-based app would enable it to reach those who have become further isolated during the pandemic and beyond [5].

This type of intervention delivery has the potential to be successfully adopted by people with dementia and their families [6], but little is known about the technological processes required to develop high-quality web apps for people with dementia and their families. However, more high-quality research is needed in this area, including more consideration of the barriers to and facilitators of use and how these impact adoption.

Aim
As part of a large research program, a paper-based manual psychosocial intervention for the PRIDE program was developed and tested for feasibility [5]. Here, we describe the processes associated with the technological work and adaptation of the manualized PRIDE intervention into a usable web-based platform, the PRIDE app. The aims of the web platform development were to (1) design an innovative log-in system tailored to the needs and abilities of people with dementia and (2) involve project stakeholders in the development of the website to ensure that the intervention is tailored to their needs, preferences, and abilities. This involvement would help involve more consideration of the barriers to and facilitators of use for the PRIDE app.

Methods

Figure 1 outlines the 2 development phases involved in creating the web-based PRIDE.
Phase 1: Preliminary Development

Overview

Work on the development of the PRIDE website began upon completion of the second draft of the PRIDE intervention [7] and ran concurrently with feasibility testing of the paper-based version of the program. The development stages of the web-based platform were (1) technological work, including project tendering and preliminary development; (2) consultations; (3) development of a beta version of the website and user testing or consultation; and (4) production of the final web platform.

Technological Work: Tendering and Preliminary Development Work

An invitation to tender was written with input from MindTech Healthcare Technology Co-operative, a National Institute for Healthcare Research–funded national center for the development, adoption, and evaluation of new technologies for mental health care and dementia. The standard university tendering procedure managed by the procurement department was followed. Developers accessed the brief, which included details of the PRIDE intervention and requirements from the web app (eg, must be user-friendly and adhere to Dementia Empowerment Engagement Programme guidelines [8]), and bid for the work contract.

In total, 26 bids were received, and 2 members of the PRIDE team independently reviewed all bids and rated them according to the standardized scoring criteria provided by procurement. The dimensions of the bids assessed included service delivery, website development, implementation plan, and data security. Each dimension received a pass or fail, and notes were made to support these ratings. A total quality score was generated based on scores from each dimension; bids were ranked, and a shortlist was made, which was reviewed by a Digital Research Specialist. The final shortlist (7 bids) was further discussed, the outcome of which was the selection of 4 software companies to be interviewed. Ayup Digital Designs was commissioned to do the work on the basis of the demonstration of an excellent understanding of the intervention, dementia-friendly design, and previous experience in health and social care–based projects.

A user-centered design approach, broadly in line with the Government Digital Service Standard Agile Delivery methodology, was used. A discover meeting attended by representatives from Ayup and the PRIDE team occurred to consolidate the company’s understanding of the intervention and discuss ideas for how the paper-based manual content and processes of the intervention would be adopted for the website. Ayup conducted work on information architecture, user journeys, user experience, and interface design. Alpha-stage wireframe designs were created and reviewed. The work outputs facilitated further discussion of how the website would work in practice (eg, how information would be navigated and presented by stakeholders).

Consultations

Multiple consultations were arranged during iterations of the website. An opportunistic sample of key stakeholders was invited to discuss the initial drafts, including log-ins, fonts, colors, and layout. Then, 3 consultations were held. The first group comprised 3 members of the University of Nottingham (UoN) Patient and Public Involvement (PPI) dementia group, which met regularly, typically attended by people with dementia, caregivers, researchers, representatives from local community organizations, and health care professionals. Participants were invited to participate in consultations following a presentation on the PRIDE project. They had not participated in any aspect of PRIDE. Members of the PPI group with dementia were actively and regularly involved in PPI, community, and research activities associated with universities and other organizations, such as the National Health Service; therefore, their participation in these consultations was not considered above and beyond their usual activities.

The second consultation involved a person with dementia, their supporters, and memory nurses who had participated in the PRIDE feasibility study. These participants had insight into the experience of receiving or delivering the intervention in practice; therefore, they could comment in depth on the content of the intervention and intervention processes and directly compare the paper-based and web app versions. The memory nurses invited dyads (people with dementia and their supporters) who had completed or were part of the intervention in the session. A third consultation was conducted via teleconference with a researcher who had delivered several intervention sessions using the paper manual and materials at a PRIDE site. The lead PRIDE researcher, who provided support and training, contacted the researcher via email with an invitation to participate.

Consultations were planned to last for a maximum of 3 hours. Examples of website wireframes (blueprints that show the basic framework of a website) were shown on a projector screen with pages adjusted for size as necessary. Before the end of the discussion, the web developer summarized the key points from the notes and asked the group to confirm if these reflected their comments. The second consultation was shorter, lasting 2 hours. A web-based videoconferencing program was used for the third, so that the wireframes could be viewed.

Discussion Topics

Discussions in the first consultation focused on (1) the use of technology to identify which devices the intervention would most likely be accessed on (eg, tablet, mobile, or laptop); (2) challenges with technology to highlight user experience; (3) the PRIDE log-in system to determine whether the innovative methods proposed were acceptable to stakeholders (easy to remember yet secure); and (4) feedback for a limited selection of wireframes and examples of design features (eg, font, color palettes, and icons) that were shown to the group.

The same discussion points were covered by participants in the second consultation, but participants were also asked how best to adapt the paper-based version of the program for delivery via the website. The group considered proposed ideas for the website presentation of activities featured in the paper-based materials (eg, completion of the profile) and how the website could be used to facilitate interaction among the person, supporter, and facilitator during the session compared with the paper-based manual and worksheets. The third consultation
focused on reviewing the wireframes and considering the functionality of the website from the perspective of an interventionist with experience in delivering PRIDE.

**Analysis**

Notes were taken at consultations by the researcher facilitating the session and the website designer. These were circulated among the teams and collated after the consultation. No formal analyses were performed on the gathered data; however, action points were generated for use by the website developer to create further versions of the website wireframes.

**Development and User Testing of the Beta Version**

Findings from user research activities were synthesized, and assumptions about user stories or website features were tested and validated. A further round of design iterations was undertaken before a beta version of the website was developed.

The beta version was reviewed by the research team and checks (e.g., spelling, grammar, and flow through the intervention process) were performed, and consultation sessions were arranged with stakeholders. The purpose of these consultations was to observe participants using the website and gather comments on usability issues such as ease of navigation. A key aspect of user testing was to enter dummy data into the activity sections of the website and set up of the log-in system.

Consultations on the beta version of the website included 4 individuals with dementia, 4 supporters, 2 PPI members, and 3 intervention facilitators. Consultations took place at the homes of consultees, in the National Health Service, or in the university departments. Researchers were provided with a topic guide, including questions, prompts, and a list of tasks for consultees to complete (e.g., logging in and out of the website). Researchers implemented a “think-aloud” protocol, encouraging consultees to comment as they used the website to yield insight into their experience, particularly areas of difficulty [9]. Comments were noted and supplemented with notes written by the researcher.

Feedback from user testing and consultations was given to the design team, who subsequently made design tweaks to the beta version to enhance usability. The full website was developed with special attention to ensure the website was as accessible as possible.

**Informed Consent and Ethics Approval**

All consultations were informal, where no personal information about the participants was collected and the discussions were not recorded. All participants provided verbal consent to participate in the discussions. Consultations were specified in the PRIDE protocol, based on which the study received ethics approval from East Midlands Nottingham Research Ethics Committee (16/EM/0044). All participants with dementia were in the early stages of the condition and were deemed to be able to provide verbal consent for their involvement by the recruiting researcher.

**Phase 2: Development of the Final PRIDE App**

**Overview**

Researchers and Ayup agreed on continuing an agile approach to app development, as it enabled dynamic collaboration between all relevant stakeholders and was also the standard practice for Ayup. As part of this approach, intensive development periods called sprints were incorporated to ensure priority work was completed within a specific timeframe. For this stage of development, each sprint would last 1 week, and Ayup’s workload would be specifically aimed at the PRIDE app.

**Initial Run-through and First Sprint**

The work on further developing the PRIDE app began in November 2019. An initial run-through of the prototype was conducted by 2 researchers at the UoN (ARL and OM), with a list of issues regarding the design, functionality, and content of the web app collated. One researcher viewed the app from a practical viewpoint, whereas the other used their knowledge and experience of working with people with dementia and viewed it from their perspective. The potential amendments were noted and discussed by the study team. A specification document was compiled and sent to Ayup, the company responsible for app development for the study. Following the initial run-through, 2 development sprints were scheduled for spring and summer 2020.

The focus for the first sprint was the highest priority issues identified with regard to the functionality, content, and overall design of the PRIDE app. Specification and priority documents were supplied to Ayup before a sprint planning meeting between the study team and development company. This provided the opportunity to discuss the workload and clarify any final improvements before the sprint start date.

**Specification and Priority Documents**

The specification document outlined the following goals and key points for the first sprint:

- Navigation to and between sections—clearer signposting of the content, such as the addition of a contents page, so users can see which section they are completing, and making the sidebar menu items more evident
- Larger font and better page layout (less empty white space)—reduction in the amount of text per page to reduce the need to scroll down the screen and increase in font size
- Addition of identifiable icons—clear and consistent use of easily recognizable icons, with particular attention given to the navigation icons including “Home,” “Help,” and “Back”
- Maintained access to introductory session content—the prototype did not allow users to revisit sessions from the first intervention session

Priority tasks were identified as fundamental, high, or low using the MoSCoW (Must Have, Should Have, Could Have, and Will Not Have this time) prioritization framework. The target was all fundamental and high-priority tasks to be completed within the first sprint. Fundamental tasks included enabling continued access to introductory session content, increasing font size, addition of show or hide tabs to reduce long sections of text, and improvements to navigation and signposting. High-priority tasks included the addition of activity icons and instructions, inclusion of a glossary link on the user’s main dashboard, and fixing graphical glitches on images. The sprint was completed
in April 2020, with all fundamental and high-priority tasks being completed. Those tasks that were of lower priority were held over to the second sprint.

User Testing
Following completion of the first sprint, it was important to obtain feedback from the target user group. Contact was made with established PPI groups at the UoN and in the local community with the aim of recruiting volunteers to provide “expert consultations.” Adverts for volunteers were posted on various social media feeds. The Alzheimer’s Society was also contacted but was unable to publicize the call for volunteers because of the COVID-19 pandemic. Two volunteers, a person living with young-onset dementia and their partner who were members of an established PPI group, were recruited for the user-testing stage. Written guidance on accessing and navigating the PRIDE app was provided, and the volunteers could contact the team if they encountered any problems. The volunteers explored the PRIDE app in their own homes over the course of a week before providing written feedback on their experiences.

Overall, the feedback gathered consisted of a mixture of positive and negative comments. The log-in process was perceived as easy to use, and the activities prompted positive discussions between users. However, they did think that some of the content was aimed at older adults with dementia, rather than young-onset dementia, and therefore might not be as relevant to those of all ages living with dementia. They also found that working slowly through each section and making notes helped people with dementia follow the content. Feedback from this stage was actively provided during the second sprint stage. Some comments from the user testing are as follows:

Logging in was straightforward.
The plan, do, review process made sense to [the person with dementia] when I worked through it with him and prompted ideas for things that would help/hinder him in the activities he wanted to try doing.
Impact of COVID on going out and socializing might need to be factored in.
Generally, [the person with dementia] found it difficult to tackle more than a few sections in one sitting. When we started work the next day, he had forgotten what he had done previously. We found working through each section slowly and making notes or drawing something to reflect our conversations made things easier.

Second Sprint
A second sprint was originally planned for the summer but owing to the difficulties in finding user-testing volunteers and the impact of the COVID-19 pandemic, the sprint was delayed until September 2020. The focus was on making the improvements and amendments identified during the user-testing stage. Similar to the first sprint, a specification document was sent to Ayup with developmental changes before the start date. For this sprint, the document highlighted grammatical errors that needed to be resolved in the content; identified words and phrases that could be changed to increase clarity and make the content more dementia-friendly; and added a paragraph regarding the impact of the COVID-19 pandemic and how this could affect their activities. This information was also uploaded to Trello, a planning software, which enabled us to prioritize actions and estimate the time taken to complete these actions. This allowed a more collaborative approach to sprint work between the study team and Ayup, and the researchers were able to monitor the progress of tasks during the sprint. All high- and medium-priority changes were made, such as correcting typographical and grammatical errors and adding a statement about how the COVID-19 pandemic could affect the ways in which people use the PRIDE app, which vastly improved the usability and functionality of the PRIDE app, bringing it up to a standard suitable for use by participants.

Before the COVID-19 pandemic, a field-testing stage was planned to follow the completion of the second sprint. Volunteers would have completed a remote run-through of the PRIDE app with the study team and provided additional feedback on the app’s usability and functionality from the perspective of the target population. However, owing to difficulties in recruitment experienced during the user-testing stage and the additional constraints and impact of the pandemic, this stage was removed.

Results
Phase 1: Preliminary Development
Overview
On the basis of the discussions of previous research on how people with dementia may use technology and their specific needs, an initial draft of the wireframes was developed. It was important for Ayup to understand the range of stakeholders’ digital literacy and the ability to best design an experience that meets their needs.

Keeping in mind the deterioration in cognitive skills characteristic of dementia, the team developed a log-in system that would not require the user to remember a password but that would uniquely identify their account and uphold security. Ayup proposed that an intervention facilitator assigned to a person with dementia create a PRIDE account for them in the first instance, which consists of basic data including name, date of birth, and contact details. This becomes their “PRIDE profile.” Once an account is created, the person can log in to the PRIDE website by entering their initials and date of birth; then, a unique, single-use 4-digit code is sent to a registered contact number via an SMS text message or an automated telephone voice message.

Log-in Process
The concept of the log-in system was discussed with consultees to determine its acceptability. Consultees with dementia in groups 1 and 2 acknowledged that dementia may affect their ability to remember passwords. They described “fear” of having passwords, feeling the information was too important to lose if forgotten, and identified potential safety risks of strategies to remember passwords such as writing them down. The idea of a log-in system using initials, date of birth, and a single-use
code was well accepted. Although the date of birth relies on memory, a consultee said that this information is a personal possession and something that never changes, so they thought it would be difficult to forget. One consultee described feeling “lucid” and able to solve problems using logic at the moment, for instance, to navigate the log-in system but that they did not know how long they would be able to do this. Consistent with this concern, a consultee in group 2 suggested the use of the proposed log-in system; there would have to be instructions on the screen to remind the person of the sequence to follow.

In contrast, intervention facilitators participating in consultations 2 and 3 suggested having log-in details saved in a browser might be a simple way to assist people to remember passwords without the need for a specific log-in system. However, people with dementia in consultations 1 and 2 were wary of saving passwords automatically through their web browser or using autofill functions, as they felt this was less secure and anyone could potentially access their personal information. When asked whether it would be preferable to receive the single-use code via telephone call or SMS text message, many consultees said that the telephone call may be a problem, as they have call-screening devices to prohibit unknown or nuisance calls. Some said that as long as they knew that they would be receiving the call, they could pick it up. The final log-in system uses a combination of initials, date of birth, and a single one-time code that is either sent via SMS text message or via telephone using text-to-speech technology.

**Paper-Based Materials Versus a Web-Based Platform**

Consultees in group 1 had not previously taken part in the PRIDE intervention but were asked if they would have a preference for paper-based or web-based materials if they were to take part in PRIDE. Two said it would be easier to use paper-based materials, adding that they did not have to think about things going wrong with technology. The person with dementia and supporters in consultation 2 who had used paper-based materials in the feasibility study preferred the website format, identifying the following benefits: (1) it would be easier for intervention facilitators to see necessary information (eg, plans) on the web rather than having to refer to several sheets of paper; (2) it would be a more effective way of delivering reminders instantly as you might forget to look at a calendar; (3) it might stimulate the person and lead to uptake of other activities such as brain training that might be helpful; and (4) it might be easier to read typed text and type text than to read and write for people with dementia.

Supporting the first point, intervention facilitators in consultation 2 added that they had experienced problems with people losing the manual and paperwork between sessions, and if the supporter was not present in sessions 2 and 3, it was difficult to determine what had actually been done without the accompanying paperwork. The intervention facilitator in consultation 3 felt the website wireframes seemed to relate to the paper-based manual quite well.

**Concerns About Technology**

Intervention facilitators stated that in their experience, many older people with dementia did not use or have computers but many used mobile phones or had computer tablets. Intervention facilitators said that the use of computers would depend on the age group and raised the point that some people may feel embarrassed or reluctant to engage if they are not computer literate. Some intervention facilitators said they were “scared” of technology but had phones and computer tablets, although they did not use them in sessions with clients. The intervention facilitator from consultation 3 said that there were participants they had delivered PRIDE to who benefited from the paper-based version of the intervention but who may not have agreed to take part if it were presented using a web-based platform, as the use of technology would be a barrier. However, they also reasoned that even if participants were not familiar with technology, they might be willing to try the right adviser. Consultees with dementia highlighted the importance of social interactions in the delivery of interventions, stating that “people should not be replaced by computers.”

Intervention facilitators raised other considerations related to technology that may disrupt delivery of the intervention, including practical issues such as the internet either not being available or working in people’s homes, paying for internet access, and the person forgetting to charge devices. However, all consultees with dementia reported using different types of technology in their daily lives to send and receive emails, search for information, watch videos, and play games, in contrast to the expectations of the intervention facilitator on computer use among this client group.

**Design and Accessibility Features**

The designs presented include samples of text, proposed website page layouts, colors, and images. The intervention facilitators participating in consultation 3 said it was important for the design and layout of the website to be simple and felt that the wireframes fulfilled this requirement well. The supporter participating in consultation 2 felt the colors needed to be brighter to make content more noticeable, commenting that “in older age eyesight isn’t as good.” They also suggested making all text, buttons, and icons that were supposed to be clicked the same color to differentiate from content without hyperlinks to other pages on the website. Some consultees with dementia had trouble identifying the meanings behind some of the images selected to represent themes, for example, a running stick figure to represent “keeping physically active.” Consultees said that black text on a white or yellow background would be the clearest to read, and certain colors carried certain meanings. For example, red is viewed as a danger. They felt that the colors presented on the wireframes were sufficiently clear. Consultees also expressed a preference for capital letters, followed by small print in text, rather than text presented in block capital. In terms of text size, consultees said that text might be too small if viewed on a mobile device and talked about the ability to change or set a particular text size on the website. Consultees felt it was a good idea to have audio-recorded versions of the text presented on the website pages for those who did not or could not read the content.
Beta Version
Following the initial consultations, Ayup iterated the website’s wireframes to incorporate several learnings. Specific developments to enhance user experience included the following:

- An option to download certain parts of the site or content for further reading offline or for printing
- An option to include a font size choice when setting up the users’ profile
- Avoiding “pop ups” that are unclear
- Changing design styles too much to keep consistency
- Prioritizing contact via a phone call when using the log-in system
- Removing block capitals and keeping all words in sentence case
- Making clickable buttons more obvious
- Placing a title next to icons so there is less ambiguity

The ability to skip through the steps in the first PRIDE session was identified as something to be modified. A linear process, by which users had to complete a sequence of 26 steps in the same order (before being able to freely navigate through the content of the website), was chosen to standardize the first session of the intervention and ensure that all compulsory activities were completed. However, the consultees felt that this made the process too lengthy and having so many steps was confusing. The intervention facilitators added that this structure might also impede their ability to tailor information to the person, which they felt was an important aspect of delivering the intervention. According to a suggestion by a consultee, “Next” buttons were added at the top of each page so that pages can be bypassed if required and “Back” buttons were added so that users can move freely between the steps according to their preference. An overall action point was to review navigation across all aspects of the website to ensure all hyperlinks connect to the correct page and refine the user journey through the “plan, do, review” content, as some of the consultees noted navigation through this information felt “circular.”

A point of frustration was that the error messages displayed when data entered into the website were not accepted or when boxes requiring data were left blank. It was not specified why the data had been rejected or which aspects of the required data on the page were missing, so it was decided that all error messages should specifically and clearly reference the issue and the location of the issue.

The PRIDE intervention manual includes a series of “case stories,” demonstrating how people have overcome the challenges associated with living with dementia. These are part of the tailored content of the intervention; thus, not all case stories will be relevant to every person receiving the intervention; rather, the intervention facilitators will select case stories they think will be helpful for the person. The intervention facilitators suggested instead of being embedded in the content of the website, which may make them difficult to locate, case stories should feature in a “narrative index,” which the intervention facilitators could refer to, to make the process of picking out examples more streamlined.

Consultees were able to easily use the log-in system. However, it was suggested that to save time, intervention facilitators should be able to register the person with an account for the website before the first session, rather than as part of the first session.

Phase 2: Development of the Final PRIDE App
Overview
The researchers discussed their consultations and agreed on a series of action points that were then provided to Ayup. The priority of amendments was negotiated using the MoSCoW prioritization framework based on the assumed importance and estimated time they would take to complete.

Through sprint work and user testing, the PRIDE app was refined and made as relevant to its target users as possible. The PRIDE app is a web-based app that is accessed through a web link rather than an app store logo. After modification and refinement, the PRIDE app became a functioning web-based handbook.

Log-in Process
Facilitators create an account for individual users using 2 initials, a date of birth, and a 4-digit code that is sent to a contact number. When users log into the app, the code is sent through either SMS text messages or voice messages, and thus is accessible to those without a mobile phone. Figure 2 shows the 3-stage log-in process.

Figure 2. The 3-stage log-in process participants use to access the Promoting Independence in Dementia (PRIDE) app.
**Introductory Session**

The session structures are the same, with advisers and users completing the same introductory session as in the paper version. After logging on for the first time, users are shown 26 different steps that constitute the session content. They can save their progress and exit the app at any time, with their next step highlighted at the top of the page when they log back. A navigation bar on the left of the screen shows the users which section they are currently in. Figures 3-5 show some of the introductory session content.

**Figure 3.** After logging on for the first time, users see the introductory session contents page. PRIDE: Promoting Independence in Dementia.

**Figure 4.** Example of the interactive activities for users to complete. The instructions were added during the sprint development.
Figure 5. Following sprint work, collapsible sections of text were created. The arrows allow users to expand and collapse the sections as they wish.

Main Dashboard
Once the introductory steps are completed, users are transferred to the PRIDE app home page interface (Figure 6). Here, they can navigate back to the introductory session, access the individual topic areas, add members to their supporter network, create further action plans, and update their activity log.

Figure 6. Promoting Independence in Dementia (PRIDE) app home page, where users can see their plans, activities, and access topic information.

PRIDE App Topics
Participants can view information on the 7 main topics included in the PRIDE app at any time. During the introduction session, users are asked to select 3 topics that they would like to focus. This selection can be amended by users at any point through the topics section on the PRIDE app (Figure 7).

Users can also use this section to learn more about each topic. There are personal stories intertwined throughout the content to provide users with insight and reassurance of how others with dementia have made positive changes across topics. Figure 8 shows an example of one of the topic pages. Users can read all of the content or access specific subsections directly.
**Creating Activity Plans**

From the Plans section, users click “+Add a plan” and select the topic for which they would like to create a plan. The topic selection given on the screen are the 3 topics users have selected to focus on. After selecting a topic, a page will appear asking users whether they would like to learn more about the topic or create a plan. Participants filled in the plan and selected whether they would like to carry on, try, do more, or do less of an activity. They can write about where they can execute this activity, facilitators, and potential barriers. Once completed, they click on “Save and submit plan” (Figure 9).
Logging an Activity
Once plans are created and saved, they appear on the user’s home dashboard. They select the plan they would like to log an activity. On the next page, they fill in what activity they completed, when, and how long the activity took. Clicking “Save and submit” will add that activity to their log on their dashboard (Figure 10).

Review Sessions
Following the introductory content, and after they have had time to use the PRIDE app in their daily lives, users completed 2 review sessions with a facilitator. From the home dashboard, users click on the begin review link and confirm that a facilitator is present. Once confirmed, the app will ask which of their plans they would like to review (Figure 11). One plan can be reviewed at a time, but the review process can be completed for as many plans as desired. The app asks users to complete boxes on how
the activity went, whether anything helped or hindered them, and what the next stages were. At the bottom of the review page, participants are asked whether they would like to leave the plan as it is, revise it, or archive it (if they are happy and feel like they have completed their plan).

Figure 11. Example of selecting a plan to review. PRIDE: Promoting Independence in Dementia.

Discussion

Principal Findings

This paper presents the development of the PRIDE app, a psychosocial intervention that targets multiple domains often affected following a dementia diagnosis. Developments to enhance the dementia friendliness of the app were achieved through collaborative sprint work and the involvement of people living with dementia. To our knowledge, the PRIDE app is unique in its content, and this is the first study to present such an intervention.

Comparison With Prior Work

A previous study on the individual Cognitive Stimulation Therapy (iCST) app helped inform our development process [10]. They took an iterative approach to app development and involved people living with dementia and their caregivers to improve the structured cognitive stimulation application. Through interviews and focus groups, the researchers were able to incorporate participant feedback into their 3 developmental sprints and explore the initial experiences of using the computerized cognitive stimulation program [10]. The iCST app was similar to the PRIDE app in that it was a one-to-one program delivered at home on a touchscreen tablet. However, the interventions differed as iCST was carer-led, only applicable to tablets, and purely focused on cognitive stimulation activities. Although there were differences, the iterative approach used was very similar to that in the PRIDE app development, as feedback from people with dementia and their supporters also informed the sprint work [10]. In addition, the discussion guide for their interviews helped inform the questions asked when gathering feedback and for the interview that will be conducted with participants following their use of the PRIDE app.

Over the last decade, there have been changes in the quantification of health and quality of life. Some have proposed an update to the World Health Organization’s definition of health by altering the focus on how well an individual can self-manage and adapt to physical, mental, and social health challenges [11]. The PRIDE app aims to meet this evidence gap by providing an intervention that covers multiple domains relevant to the revised World Health Organization definition and targets a range of self-management concepts.

A previous systematic review of web- and app-based interventions for dementia showed their potential to produce positive outcomes in self-management and can be successfully delivered through a range of methods [6]. Existing interventions targeted several self-management concepts, such as independence and activities of daily living, but there was an inconsistency in which domains often affected by dementia were targeted by interventions, and some purely focused on 1 concept. The review also revealed a lack of high-quality evidence on these types of dementia interventions and no studies researching an intervention that encompasses physical, cognitive, social, and emotional domains. The PRIDE app aims to meet this evidence gap by providing an intervention that covers multiple domains and targets a range of self-management concepts.

Limitations

We did not foresee the low recruitment of user-testing volunteers, and this delay had a wider impact on the study timelines. Despite the call for volunteers to go out to local and national groups, there was very limited interest in user testing.
This was likely because of the COVID-19 pandemic, the change in people’s priorities, and lack of interest in research. However, the recruited volunteers were experienced in dementia studies and provided useful feedback. Another limitation is the removal of the field-testing phase. Originally, this stage was to be incorporated following the second sprint to assess the PRIDE app’s usability and accessibility, with a third sprint proposed to resolve any urgent problems. A delayed field-testing phase was not a viable option for this study because of the time constraints and resources available for the study.

The COVID-19 pandemic meant that remote working was necessary for the vast majority of the PRIDE app development. It also required community and PPI groups for people living with dementia and their families to either close temporarily or move on the web. Unfortunately, these required measures contributed to difficulties in finding user-testing volunteers and removing an accessible source of feedback during the ongoing app development. Delays caused by these difficulties led to the second sprint, which had a knock-on effect on the rest of the study timelines. As diagnoses were not recorded from those involved in the development of the app, conclusions about specific types of dementia were limited. This should be considered when conducting future research to understand any potential barriers specific dementias could cause.

Following the development work, the PRIDE app will be the focus of the reach, effectiveness, adoption, implementation, and maintenance (RE-AIM) study [12]. The app offers people living with dementia a central source of information and support on a range of domains commonly affected by dementia; and this study will explore the potential reach, effectiveness, and adoption of the intervention. Although a larger trial will be needed to assess the potential effectiveness more comprehensively, the RE-AIM study will provide initial insight into whether the PRIDE app could be a feasible intervention, suitable for further research, and whether it could have positive outcomes for people with dementia and their families.

**Conclusions**

The PRIDE app has evolved from its initial prototype [5] into a more dementia-friendly and usable program that is of a standard suitable for wider testing. It has the potential to advance the previous evidence into web- and app-based interventions, in addition to providing better support for self-management, improving individuals’ level of independence, and enhancing the quality of life of people with dementia and their families. The finished version will be tested in a RE-AIM study, with its potential reach, effectiveness, and adoption explored. This study will contribute further to the evidence base and our understanding of how web- and app-based interventions could be successfully implemented in dementia management. Feedback gathered during the RE-AIM study will lead to any further developments to the app to increase its applicability and usability to the target audience, such as considering alternative log-in methods and identifying barriers for specific dementia types. It will also provide further understanding of the barriers and facilitators that have a significant impact on the adoption of these interventions and how these could be overcome in future research.

**Acknowledgments**

The National Institute for Health Research Applied Research Collaboration East Midlands supported this study. The views expressed are those of the authors and not necessarily those of the National Institute for Health Research or Department of Health and Social Care.

**Authors’ Contributions**

ARL, EC, LY, EM-C, OM, DK, and MO drafted the manuscript. The conception and design of the Promoting Independence in Dementia program were overseen by EC, LY, EM-C, and MO, with ARL, OM, ST, and MS contributing to the development and production of the web app. Data collection was conducted by ARL, EC, LY, DK, ST, and MS.

**Conflicts of Interest**

None declared.

**References**


Abbreviations

iCST: individual Cognitive Stimulation Therapy
MoSCoW: Must Have, Should Have, Could Have, and Will Not Have this time
PPI: Patient and Public Involvement
PRIDE: Promoting Independence in Dementia
RE-AIM: reach, effectiveness, adoption, implementation, and maintenance
UoN: University of Nottingham
Web-Based COVID-19 Dashboards and Trackers in the United States: Survey Study

Melissa D Clarkson¹, MDes, PhD
Division of Biomedical Informatics, University of Kentucky, Lexington, KY, United States

Corresponding Author:
Melissa D Clarkson, MDes, PhD
Division of Biomedical Informatics
University of Kentucky
725 Rose Street
Lexington, KY, 40536
United States
Phone: 1 859 323 7232
Email: mclarkson@uky.edu

Abstract

Background: The SARS-CoV-2 pandemic provided an opportunity to use public-facing web data visualization tools to help citizens understand the evolving status of the outbreak. Given the heterogeneity of data sources, developers, tools, and designs used in this effort, it raised questions about how visualizations were constructed during a time when daily batches of data were available, but issues of data quality and standardization were unresolved.

Objective: This paper surveyed web-based COVID-19 dashboards and trackers that are likely to be used by the residents of the United States to monitor the spread of infection on a local, national, and global scale. This study is intended to provide insights that will help application developers increase the usefulness, transparency, and trustworthiness of dashboards and trackers for public health data in the future.

Methods: Websites of coronavirus dashboards and trackers were identified in August 2020 using the Google search engine. They were examined to determine the data sources used, types of data presented, types of data visualizations, characteristics of the visualizations, and issues with messy data. The websites were surveyed 3 more times for changes in design and data sources with the final survey conducted in June 2022. Themes were developed to highlight the issues concerning challenges in presenting COVID-19 data and techniques of effective visualization.

Results: In total, 111 websites were identified and examined (84 state focused, 11 nationwide, and 16 with global data), and this study found an additional 17 websites providing access to the state vaccination data. This study documents how data aggregators have played a central role in making data accessible to visualization developers. The designs of dashboards and tracker visualizations vary in type and quality, with some well-designed displays supporting the interpretation of the data and others obscuring the meaning of the data and potentially misleading the viewers. Five themes were identified to describe challenges in presenting COVID-19 data and techniques of effective visualization.

Conclusions: This analysis reveals the extent to which dashboards and trackers informing the American public about the COVID-19 pandemic relied on an ad hoc pipeline of data sources and data aggregators. The dashboards and trackers identified in this survey offer an opportunity to compare different approaches for the display of similar data.

Introduction

Background

SARS-CoV-2, a novel coronavirus, was first detected in the United States in mid-January 2020 [1,2], and eventually, many states enacted stay-at-home orders in early March. The SARS-CoV-2 pandemic challenged the public health system in the United States in many ways, including a lack of laboratory testing capacity early in the pandemic, evolving data standards for reporting positive test results and deaths owing to COVID-19, and a lack of coordination among state and federal...
agencies. In addition, approximately half of all nonfederal hospitals lacked the capacity to electronically exchange information with public health agencies at the beginning of the pandemic [3].

The pandemic also presented challenges in communicating with the public the evolving status of the outbreak and the reasoning behind public health measures, such as stay-at-home orders and masking. As the pandemic progressed, waves of infection rose and fell in the regions of the United States at different times owing to the presence of superspreader events, differences in public health responses among states, and the rise of variants [4,5].

Data Visualization

Data visualization has the potential to modify the course of a pandemic by bringing together information about the state of the pandemic, public policy, and individual behavior in ways that are actionable. However, for the visualizations to have this impact, they must be easily accessible, based on accurate and timely data, and carefully developed with an understanding of both the data and the principles of visual design.

Visualizations have been easily accessible during the pandemic owing to the availability of numerous software tools and platforms for creating graphics and mapping data. Because data sets are available to anyone with an internet connection, early in the pandemic, a number of visualization experts wrote about the need to responsibly use tools and data when creating visualizations [6-9]. Misrepresented and misinterpreted COVID-19 visualizations have inspired one study to use them to help students develop statistical literacy [10].

The large number of visualizations developed and deployed rapidly by public health authorities and data analysts during the pandemic is of interest to visualization and communication researchers. They provide insights and lessons about the process of rapidly designing and developing visualizations [11-15]; efforts to curate global data [16]; the types of visualizations created and who they are for [17-19]; conceptual models linking tools, data, visualizations, and users [20]; and what it means for a visualization to be actionable [21]. In addition, studies have used pandemic data to understand how the users perceive the risk and severity of the pandemic [22-24] and their reactions to the designs of dashboards [25].

Purpose of This Study

This study complements earlier work by taking a US-focused look at COVID-19 dashboards from August 2020 to June 2022. It documents data sources and data aggregation efforts, identifies themes relevant to designing dashboards for outbreaks, and highlights issues with data availability and standardization. The goals of this work were to provide insights that will help application developers increase the usefulness, transparency, and trustworthiness of dashboards and trackers for public health data in the future and to document the variety of dashboards and trackers used by the residents of the United States and the evolution of these tools for approximately 2 years.

This study encompassed the following 2 broad categories of data visualization: dashboards and trackers. The term dashboard generally refers to a set of dynamically updated data visualizations placed in proximity to one another and is used to monitor conditions for the purpose of understanding a system or event. Because several COVID-19 data visualizations took other forms, such as visualizations arranged sequentially with accompanying text, I used the term tracker to more broadly refer to these types of dynamically updated displays.

Methods

This survey began in August 2020, which was approximately 5 months after stay-at-home orders generated widespread public interest in the state of the pandemic and all states were providing data on the web about the pandemic for a public audience.

Identification of Dashboards and Tracker Websites

To identify web-based dashboards and trackers, I performed a web search using Google on August 12, 2020. Searches were formatted as “coronavirus COVID dashboard tracker” combined with either a state name, “United States,” or “global.” The first 15 results for each keyword combination were examined for their relevance. On the basis of test searches, I determined that the relevant search results were generally in the first 10 results, and results ranked lower than the 15th search result were either links to the dashboards and trackers from other web pages or news or commentary about the pandemic. The websites that were determined to be relevant to this study are listed in Multimedia Appendix 1. Dashboards and trackers were categorized as state focused, nationwide, or global. This study focused on visualizations that are the most likely to be viewed by people in the United States to understand the local and regional status of the pandemic, with less emphasis on global visualizations.

By the end of January 2021, many states had incorporated vaccine dashboards into their state dashboards. To locate vaccine dashboards not integrated into state websites, I performed a Google search for “covid vaccine dashboard tracker” and examined the first 20 results for relevance. Multimedia Appendix 2 lists the websites determined to be relevant to this study.

Inclusion and Exclusion Criteria

Websites were included in this survey if they displayed up-to-date information, appeared to be updated daily (or nearly daily), and relied mainly on graphs or maps (rather than tables or text) to convey information. Websites were excluded if they showed data limited to regions smaller than a state (such as a single county or city), were specific to a type of setting (such as prisons), or displayed only trackers or dashboards that had been embedded from other websites. The District of Columbia was included in this survey, but the US territories were not. This survey included only publicly available websites accessible on a laptop. Apps developed specifically for smartphones were not included.

The focus of this study was the display of information concerning diagnosed cases of COVID-19, deaths attributed to COVID-19, testing for COVID-19, and vaccination. Visualizations of risk levels, hospital bed availability, and
hospital admissions were not central to this study, but designs for these types of data may benefit from these findings.

**Methods of Review**

All websites were examined on a MacBook Pro laptop (Mac OS version 10.11) using a Firefox web browser. The data sources used by each dashboard or tracker were documented based on statements from the website itself [26], and for one dashboard, the development team was contacted. Some websites included a statement stating that their data sources have changed as the pandemic has developed, suggesting that their list of sources may not be complete or current. The software tool or method used to create visualizations was also determined. If the name of the software brand was not displayed with the visualization, the Inspector tool within Firefox was used to examine the webpage’s HTML and determine the tool or method used.

Each website was examined to determine the following:

1. Does the website credit a data source or sources?
   a. What sources are credited? How are they credited?
   b. When more than one data source is credited, is it clear which measures come from which source?
   • Rationale: Citing data sources increases the trustworthiness of visualizations; however, there is no established best practice for how to do this. Listing the name of an organization that provided the data may not be sufficient if the data set from that organization cannot be identified with certainty. However, members of the public may not expect data sources to be cited.

2. What types of data are presented?
   a. What measures are provided? (such as number of cases, number of tests performed, and number of hospitalizations)
   b. What is the level of granularity? (county level or state level)
   • Rationale: Many different measurements relating to COVID-19 were collected by different organizations and public health authorities, with new measures introduced and others discontinued. Differences in granularity are important both for describing the pandemic with more precision and for making the data more relevant to viewers (who have an interest in knowing about COVID-19 in their own area).

3. What graphical forms of visualizations are used? (bar charts, line charts, choropleth maps, etc)
   • Rationale: Surveying graphical forms provides information on which forms designers believe are appropriate for public-facing visualizations and the variety of forms available in visualization tools.

4. Do the visualizations clearly display the data? Might any visualizations lead viewers to make inappropriate conclusions?
   • Rationale: Drawing on my experience as an information designer and instructor for a data visualization course, I examined the designs for issues involving color, size, and labeling; misleading use of space or positioning; and mismatches between the type of data and the chosen graphical form. These present opportunities to increase awareness of good design in data visualization.

5. How do the designers deal with messy data, such as lags in reporting and discontinuities in definitions of measures?
   • Rationale: Identifying effective methods for accommodating messy data will help establish best practices.

**Capturing Changes in the Design of Visualizations Over Time**

To understand how the dashboards and trackers evolved over the course of the pandemic, the survey of websites was repeated 3 more times. This survey spanned from approximately 7 months after the novel coronavirus was first detected in the United States to nearly 2-and-a-half years after detection. The second review of each website was conducted between January 2021 and March 2021. The third review of each website was conducted in either December 2021 or January 2022. The final review was conducted in June 2022. By the end of the survey period, >1 million deaths in the United States were attributed to COVID-19, with deaths decreasing to <400 per day by June 2022.

**Results**

**Dashboards and Trackers Identified**

**State Focused**

A total of 84 dashboards and trackers focusing on COVID-19 cases in a state (or region composed of several states) were identified. These are listed in Multimedia Appendix 1, with each assigned an identifier in the format S-x. (This paper will refer to dashboards and trackers using square brackets with the identifier from the appendix, for example, [S-1] for the dashboard from the Alabama Department of Public Health). At least one dashboard or tracker website was provided by the public health authorities in each state and the District of Columbia as of August 2020. The Massachusetts Department of Public Health originally provided only a downloadable PDF document before switching to a dashboard created with Tableau. An additional 20 dashboards and trackers were developed by newspapers and television news organizations. The remaining websites were associated with nonprofit organizations (n=2), web-based media and marketing companies (n=2), individuals (n=2), a university-associated team, and a health care–related trade organization. All state-focused websites identified in this...
survey provided data at the county or parish level. As of June 2022, a total of 23 of the 84 dashboards and trackers were removed or no longer updated with new data. Of these, the Florida Department of Health discontinued its dashboard but replaced it with weekly reports that could be downloaded as PDF documents.

**Nationwide Coverage**

In total, 11 websites that displayed data for the entire United States were identified. These are listed in Multimedia Appendix 1, with identifiers in the format N-x. Of these, 5 websites displayed data at the state level, whereas 6 provided more granular data at the county level. The Centers for Disease Control and Prevention (CDC) provided 2 trackers [N-1 and N-2]. Other websites were provided by news organizations (n=3), university-associated teams (n=3), technology or web companies (n=2), and a nonprofit organization. As of June 2022, a total of 2 of the 11 websites were discontinued.

**Global Coverage**

An additional 16 websites were identified that displayed worldwide COVID-19 data. These are listed in Multimedia Appendix 1, with identifiers in the format G-x. These websites were provided by news organizations (n=3), university-associated teams (n=4), nonprofit organizations (n=3), and technology or web-based businesses (n=6). As of June 2022, a total of 4 of the 16 websites were removed or no longer updated.

**Vaccine Distribution**

Multimedia Appendix 2 lists the additional dashboards and trackers for vaccine distribution. This survey identified 17 state-focused sites with county-level data, 4 with nationwide coverage at the state level, and 3 with global coverage.

**Visualization Tools and Methods**

The most popular software platforms used for state-focused dashboards and trackers, particularly among public health authorities, were Tableau, ArcGIS, and Microsoft Power BI. Some dashboards presented all the information in a single page, but it was common for dashboards to have multiple pages to accommodate maps and new types of data that became available during the pandemic. News organizations were more likely to provide trackers arranged as a series of data visualizations with textual explanations and use scalable vector graphics embedded in their web pages. See Multimedia Appendix 1 for information on the visualization tools or methods used for each dashboard and tracker.

**Data Sources and Data Aggregators**

**Overview**

As with all data visualizations, it is important for the viewers of COVID-19 dashboards and trackers to know the data sources. A visualization could display data collected by the organization that created the visualization (in the case of public health authorities), data obtained directly from one or more public health authorities, or data from a data aggregator service. Multimedia Appendix 1 documents the data sources stated on the websites. Multimedia Appendix 3 provides a list of data aggregators and prominent dashboard developers with URLs for details on their methodologies and data sources.

**State Focused**

None of the websites provided by state-level public health authorities provided details about data sources or methodology, but it is likely that the data were submitted by local public health departments that received reports from diagnostic laboratories, health clinics, and hospitals. Of the nongovernmental state-focused websites, most stated that the data were from the state public health authority (or, in some cases, a combination of state and local public health authorities), but it is unclear whether these websites were drawing data directly from the public health authorities they credited or if they used a data stream from a data aggregator service. Two nongovernmental websites did not state a source of data or removed the statement [S-30] and [S-54]. One website’s data source [S-77] was credited to a data aggregator.

**Nationwide Coverage**

Throughout the fall of 2020, the CDC provided only state-level COVID-19 case counts to the public rather than county-level data. Therefore, any website displaying county-level case counts for the United States relied on data aggregated from local and state sources by a nongovernmental data aggregator. Figure 1 shows the major data aggregation pathways for case counts and testing data for the United States as of August 2020. It was created by examining data sources and methodology information for the websites and consulting additional reports [26,27]. The following are the 4 major data aggregators used to independently aggregate nationwide data:

- **USAFacts:** A nonprofit civic initiative that gathers government data [28]. County-level data available for download.
- **IPoint3Acres (CovidNet):** A volunteer group founded by first-generation Chinese immigrants in the United States [15,29]. County-level data available for download.
- **The New York Times:** County-level data available for download.
- **The COVID Tracking Project:** A volunteer organization launched by The Atlantic [30]. State-level data available for download or through an application programming interface (API). It includes data for case counts and total number of tests. This project ended in March 2021, one year after it began.

As shown in Figure 1, in August 2020, only state-level data and not county-level data were available to developers by API. During this survey, several additional resources that claimed to provide APIs for county-level data scraped from the websites of data aggregators were noticed; however, this was in violation of the terms of the service set by those data aggregators. No nationwide website appeared to use CDC as their only data source. Instead, websites relied on an independent data aggregator or a combination of use of data from the CDC and a data aggregator. Of particular interest is that the county-level tracker provided by the CDC [N-2] credited the USAFacts aggregator as its source of county-level data. In August 2020, a footnote stated “Data courtesy of USAFacts.org downloaded...
each day at 4:00 pm EST or when earliest update is available” [31]. The web page redesign in November 2020 provided more extensive details on data sources, including the statement “The COVID-19 case and death metrics are generated using data from USAFacts that CDC modifies.” The use of USAFacts was later discontinued and county-level data were obtained directly from the states [32].

Figure 1. Major data aggregation pathways for the United States’ cases and testing data as of August 2020. References in blue correspond to dashboards and trackers. API: application programming interface; CDC: Centers for Disease Control and Prevention; CSSE: Center for Systems Science and Engineering; G: global; N: nationwide.

Global Coverage

The dashboard developed by the Johns Hopkins University (JHU) Centers for Civic Impact displays data from the JHU Center for Systems Science and Engineering (CSSE). JHU CSSE acts as an aggregator of aggregators for worldwide data, relying on a large number of sources, including The COVID Tracking Project and 1Point3Acres for the US data [26]. The complete list of data sources used by the JHU CSSE since January 2020 is provided in their data repository [33].

Issues in Trust and Transparency

Trust and transparency are emphasized in the guidelines the World Health Organization has assembled for communicating with the public about disease outbreaks [34]. Dashboards and trackers may inform viewers of visualizations about the sources of the data in several ways. The most direct approach is to provide the data source within a caption for each map or graph; however, this may not be feasible for dashboards combining several visualizations. Websites using data aggregators often simply state one or more sources for the entire collection of visualizations. One nationwide dashboard [N-11] was particularly vague about the relationship between the visualizations displayed and the sources of data, crediting CDC, WHO, The New York Times, JHU, Corona Data Scraper, and official state and county health agencies without providing further details. When websites list sources in this manner, it raises the following questions:

- Is this website using a data aggregator, but crediting the sources used by the aggregator rather than the aggregator?
- Which measures from which data sources are used in a particular visualization?
- Are all these sources currently used, or is this a list of all sources ever used?
- If only one organization is listed, what is the specific data set from the organization that was used?

Early in the pandemic, data scientists raised concerns about the quality of COVID-19 data [35,36]. The challenges of collecting global data appropriate for display and analysis have led to questions regarding the methodologies and sources used by some aggregators. For example, Worldometer is a private company known for its web counters that estimate world statistics. It became an aggregator of COVID-19 data and provider of popular COVID-19 trackers [G-12] and has been criticized for having an anonymous curation team and opaque methodology [37].

Visualization Tools and Methods

Multimedia Appendix 1 presents the tools and methods used to construct the visualizations examined in this survey. With the exception of Massachusetts, all state public health authorities provided a web page displaying a dashboard or tracker in August 2020 (with Massachusetts providing PDF downloads). Websites of state public health authorities were often constructed using ArcGIS, whereas state-focused websites from other types of developers relied on a variety of tools (including Tableau, Datawrapper, Infogram, and Microsoft Power BI). Websites
providing nationwide coverage tended to be constructed with frameworks using embedded scalable vector graphics. Global dashboards and trackers were created with a variety of methods.

Critique of Visualizations

Overview

Overall, 5 themes were identified from the data visualizations and designs of the dashboards and trackers. Multimedia Appendices 4-18 provide screenshots of the websites taken during the 4 rounds of review. Not every page of the website was captured for multipage websites, but the most relevant visualizations are documented.

Theme 1: Data as Imperfect Representation of Reality

Although the data presented in COVID-19 visualizations are intended to reflect the state of the pandemic, Figure 2 provides examples in which short-term patterns and trends are owing to the methods of data collection and reporting.

Figure 2. (A and B) Display of daily deaths in California through August 2020 in 2 different dashboards. Notice line for 14-day rolling average in example B. Example A is from [S-9] on August 20, 2020. Example B is from [S-8] on August 29, 2020. (C) Example of 7-day rolling average for daily vaccinations in Illinois. Low values on weekends likely reflect delays in data reporting. From [Svac-6] on January 31, 2021. (D) Tests per day in Indiana. Notice the gray box marking preliminary data. From [S-29] on June 12, 2022. (E) Deaths per day in Colorado, showing a spike on April 24 owing to the inclusion of probable deaths. From [S-13] on August 29, 2020. (F) Cumulative deaths in Colorado, showing a dip on May 15 owing to the change of definition to include only patients who are recorded as dying from COVID-19, rather than testing positive at time of death. From [S-13] on August 14, 2020.

Theme 1a: Data as imperfect representation of reality

Theme 1b: Inconsistent definitions in data collection and reporting
Theme 1a—Temporal Data Reflect a Combination of Reporting Activity and Public Health Reality

Short-term trends in the data organized by the date of reporting can be misleading. As explained by The COVID Tracking Project:

...this data displays very strong day-of-week effects and is also extremely vulnerable to predictable rise-and-drop artifacts after holidays or other major disruptions, like storms and natural disasters, that affect the ability of counties and states to report their data. [38]

To help viewers disregard day-of-the-week variability, most time series graphs include 3-, 7-, or 14-day rolling averages. As time series graphs will have incomplete data for the most recent days (owing to a lag in reporting), the best designs visually indicate the span of incomplete data.

Theme 1b—Inconsistent Definitions in Data Collection and Reporting

In the United States, much of the public health infrastructure is regulated and managed at the state and local levels. Therefore, states have different processes for collecting data and use inconsistent definitions. For example, states vary in how they define deaths attributable to COVID-19, whether the number of tests (and positive and negative results) reflects unique people or number of specimens [39], and the diagnosis of asymptomatic cases [40]. In the early months of the pandemic, several states combined the counts of polymerase chain reaction tests (a diagnostic test) and antibody tests (which detect an immune response), leading to distortions in the data on infection rates and testing capacity [41,42]. If data aggregators were unaware of this heterogeneity in state-level data, or unable to correct for known differences, visualizations that provide state-to-state comparisons will be inaccurate. In addition, some states have reported a count of recovered patients with COVID-19. Not only did these states use different definitions for recovered, but referring to patients as recovered when the long-term effects of COVID-19 are not known is misleading [43]. Another potential source of confusion occurred later in the pandemic as people became reinfected, meaning that case counts no longer represented unique individuals if states followed the national case definition [44,45]. The Iowa Department of Public Health noted this change with the following statement:

On September 1, 2021, IDPH adopted the updated 2021 COVID-19 national case definition. As part of this case definition, IDPH began including in its total case counts individuals who were previously reported as a confirmed or probable case, but have become infected again. [S-28]

Data regarding vaccinations also had inconsistencies early in 2021. As explained by the Washington Post in a footnote below the graphs of state vaccination doses administered by day:

Data before Jan. 12 is inconsistent. On Feb. 19, the CDC altered its reporting of doses administered by federal agencies by adding them to the states where the shots had been given. From Feb. 23 forward, the data reflects doses administered to residents of the states rather than doses administered by the state. [Nvac-3]

Theme 2: The Importance of Context for Interpretation

Data require context for interpretation, and therefore, data visualizations should provide context to help viewers find meaning in a visualization. Figure 3 shows successful and unsuccessful examples of providing context.
Figure 3. (A) Comparison of vaccination rates in counties of Minnesota against state average. From [Svac-10] on September 25, 2022. (B) Timeline of waves of new cases in California compared with other states. From [S-10] on January 31, 2021. (C) Comparison of vaccination rates in Wisconsin within demographic categories. From [Svac-17] on January 3, 2022. (D) Comparison of vaccination rates in Minnesota within race and ethnicity categories. Graphing on a scale of 100% of the population (rather than proportional to race and ethnicity) makes this design less effective than example C. From [Svac-10] on January 3, 2022. (E) Comparison of percent vaccinated with 1 dose and 2 doses against the eligible population and total population of Florida. From [Nvac-3] on April 30, 2021. (F) Indication of the shelter in place policy as gray band with time series data showing newly reported hospital cases in Georgia. From [S-23] on January 31, 2021. (G) Time points for policy decisions to open or restrict public gathering in Alabama, with time series data showing reported cases. From [N-7] on December 11, 2021.

Theme 2a—Supporting Meaningful Comparisons
Many types of interpretations rely on comparisons. In the context of COVID-19, useful comparisons include differences between regions, differences between demographic groups, differences over time, and differences between vaccinated and unvaccinated populations. It is these comparisons that give meaning to the data.

Theme 2b—Indicating Changes to Public Health Policy in Time Series Visualizations
Public health policy affected the trajectory of the pandemic, and policies varied at the state, county, and city levels. Several visualizations superimposed policy changes over time series data.
Theme 3: Choosing Values to Display

Overview

The COVID-19 pandemic has provided web application developers with access to data and public interest in the visualizations of these data. However, creating useful visualizations often requires more than graphing the raw numbers supplied in a data stream. The examples in Figure 4 demonstrate why it is important to consider whether it is most useful to display the data directly as obtained, a transformation of the data, or cumulative values.

Figure 4. (A) Graphs of the 3-day average of cases (upper graph) and the cumulative number of cases (lower graph) in Colorado. Note that the decrease in new cases in June is difficult to detect in the cumulative graph. From [S-12] on August 15, 2020. (B) Cumulative number of cases in Maine. From [S-37] on August 15, 2020. (C) Cumulative number of persons by vaccination status in Hawaii. The category initiating refers to the first dose, completing indicates receiving both the first and second dose. From [S-25] on February 27, 2021. (D) Case counts by age group (upper graph) and case rates by age group (lower graph) in Michigan. The lower graph shows that patients aged ≥80 years have a higher case rate than the other groups. From [S-40] on August 15, 2020. (E) The home page of the Florida dashboard, with a map showing case counts per county. A note at the bottom says “Comparison of counties is not possible because case data are not adjusted by population.” A color-coding key was not provided. A map displaying the rates by county is available on another tab. From [S-20] on August 15, 2020. (F) Map showing case counts by county. A color-coding key was not provided, but the intensity of red reflects areas of higher population density (with the location of universities indicated). From [S-17] on August 20, 2020. (G) Case counts per county for New York City Long Island has the highest number of cases but also the highest population density.

Theme 3a—Limited Usefulness of Cumulative Counts

Many dashboards state the total number of COVID-19 cases and deaths, and some also display a time series of cumulative counts. The total number of deaths may be of general interest, but graphs of the cumulative number of cases or deaths are less useful because they show only a rising curve without clearly showing trends during the pandemic. However, it may be that showing a time series of the cumulative number of vaccinated people in a region could help persuade others to become vaccinated.

Theme 3b: Total counts are less informative than population-based rates

https://humanfactors.jmir.org/2023/1/e43819
Theme 3b—Total Counts Are Less Informative Than Population-Based Rates

The availability of county-level data helps viewers to understand the geographic distribution of COVID-19 cases and deaths. However, to be more meaningful, data should be displayed as rates (eg, number of cases per 100,000 people) rather than as counts. Visualizing count data on a map is likely to simply show areas with a higher population density and give a misleading impression that COVID-19 has not affected rural areas.

Theme 4: Choosing the Graphical Form of the Visualization

The graphical forms of the visualizations (including line charts, bar charts, and choropleth maps) and how they were arranged in the dashboards revealed a mixture of effective designs that made good use of perceptual principles as well as less effective designs. Examples are shown in Figure 5.

Figure 5. (A–C) Summary visualizations provided by the Centers for Disease Control and Prevention at the top of their COVID Data Tracker web page [N-2]. Example A was captured on February 1, 2021. This design uses pink shading below the line indicating “Cases in US, last 30 days.” This is misleading because the height of the shading does not begin at 0. Example B is the same design captured on February 26, 2021 that deceptively implies that cases have dropped to 0. Example C is the revised design of the summary visualizations captured on December 11, 2021. The shading has been removed and an arrowhead is added. (D) Top-of-page summary provided by the Denver Post [S-14]. The design allows viewers to quickly see and compare trends. Shows data from last 3 months but not from the last 2 days. Captured on August 15, 2020. (E) The first 4 columns and 5 rows of a table comparing each state, ordered by case rates. Sparklines indicate trends over time, but the span of time shown is not defined. Orange bars represent current case rates. From [S-84] on December 11, 2021. (F) Summary for Oregon. This example is less successful in communicating trends because rolling averages are not used. From [S-59] on August 15, 2020. (G) A small portion of a state-by-state comparison provided by the New York Times [N-4] using a small multiples layout. Captured on October 4, 2022. (H) Stacked time series comparing confirmed cases and probable cases in Massachusetts. From [S-39] on January 31, 2021. (I and J) Graphs comparing tests administered and test positivity rate using dual axis graphs. This design is more difficult to interpret than stacked time series. Example I is Colorado data from [S-14] on March 20, 2021. Example J is California data from [S-8] on August 15, 2020.
Theme 4a—Simple Graphs for Overview and Comparison
One challenge is to distill the data into simple but meaningful visualizations. Several websites offered simple summary graphics, often in the form of simple time series graphs or sparklines, to communicate the trajectory of the pandemic. However, these simple overviews are only effective if the rolling averages are displayed. Because the pandemic was not uniform across the United States, visualizations also helped people compare the current status and trajectories of different states. However, the key to making these comparisons meaningful is that the underlying data must be comparable, and this relies on the uniformity in data collection or adjustments by data aggregators.

Theme 4b—Comparing Different Data Sets Over the Same Timespan
Data displayed as time series are crucial for communicating about the pandemic, and meaning is often derived from comparing different types of data or data from different regions. Small multiples and stacked time series were effective in aiding comparisons. A number of dashboards provided dual axis graphs, often for comparing the numbers of coronavirus tests administered and the positivity rates over time. However, this dual axis design is difficult to interpret, and alternative designs provide better solutions [46,47].

Theme 4c—Interactivity of Graphs
Frameworks for developing web visualizations often include functionality for displaying the values of data points when the cursor hovers over points. This method of providing details-on-demand is useful for enabling an in-depth exploration of graphs [48] and is often used in time series. Another type of interactivity is to enable a viewer to customize a graph by controlling the data or presentation style through drop-down menus or radio buttons. In this survey, I noted options for choosing between case counts and case rates, setting the length of time for a time series, filtering by demographic group, and switching between a linear or logarithmic scale for case counts. When display options are provided, it is important that a default display is chosen that is suitable for the greatest number of users and minimizes misinterpretation. For example, a linear scale should be the default, but advanced users may choose the option of a logarithmic scale [49]. One particularly useful option for understanding the global spread of the coronavirus is to align outbreaks in different countries based on days since a country’s outbreak reached a particular threshold of cases rather than by date. The former option is the default for a graph provided by Our World in Data [G-10].

Theme 5: Pitfalls of Automated Data Display
Overview
Dashboards and trackers visualize streams of data that are automatically updated. This combination of dynamic data and the lack of human oversight revealed some pitfalls that should be avoided to build more robust systems. These findings also suggest that dashboards need frequent monitoring to detect problems in the design of displays or the handling of data. Figure 6 demonstrates several of the identified problems.

Theme 5a—Display of Peculiar Data
Some anomalies in the displayed data cannot be explained by small adjustments to the data or artifacts such as day-of-the-week variations. Extremely high or negative values of counts indicate problems in recording, processing, or transmitting data. The presence of these anomalies should alert developers (and viewers) that the trustworthiness of the entire data set and visualization is questionable.

Theme 5b—Designs May Cease to Support Meaningful Comparisons
A design that works well with a particular range of values or size of data set may lose effectiveness as data are dynamically updated. For example, a method of binning data that is effective early in the pandemic will become much less informative if all the data are represented within a single bin later in the pandemic. However, one drawback of adjusting bins over time is that people who periodically view a graph may assume that changes in the distribution of data in bins reflect changes in the data rather than in the definition of the bins.
**Figure 6.** (A) Deaths per day for Colorado, including peaks of −170 and 841. From [N-10] on April 10, 2021. (B) Newly hospitalized patients per day for Kansas, including peaks of 5417, 7257, −9387, and −5290. From [N-10] on April 10, 2021. (C and D) Color coding of counties in Michigan based on case rate. Captured on August 15, 2020, and March 20, 2021. By March 2021 all counties are in the highest bin. From [S-41]. (E and F) Alternate view of Michigan map captured on the same days that display total case counts (rather than rate). Uses same color-coding key as examples C and D. Notice that the data on the August 2020 map spans 5 bins, whereas the March 2021 map uses only 3. (G and H) Patient status in Oklahoma. Captured on August 15, 2020, and April 10, 2021. By April, the number of recovered cases make the length of the active case bar unreadable. From [S-58].

**Theme 5a: Display of peculiar data**

**Theme 5b: Designs may cease to support meaningful comparisons**
Discussion

Principal Findings

This study identified and examined >100 websites providing COVID-19 dashboards and trackers relevant to the residents of the United States and highlighted the multitude of factors that affect these visualizations. The findings reveal the role data aggregators have played in making data accessible to visualization developers as well as lapses in communicating to viewers the provenance of the data. Decisions by public health experts about data collection and data standards have downstream effects on which data are available to be communicated and compared. In addition, each step of this process is impacted by the evolving nature of the pandemic and political and social systems.

The five themes identified in this work can guide future development of visualizations of public health data for the public: (1) viewers should be made aware that data are an imperfect representation of reality owing to methods of data collection and reporting; (2) viewers need context for interpreting visualizations, such as comparisons with other data or indicators of relevant events on timelines; (3) developers should carefully consider whether plotting a raw data stream, cumulative values, or transformation of values will be the most useful to viewers; (4) the graphical form of a visualization should be chosen to fit the type of data and be designed to make good use of perceptual principles; and (5) visualizations designed to use automated streams of data must be monitored to ensure that the data continue to have reasonable values and that the design of the visualization remains useful with the new data.

Trust and Transparency Begins With the Data

One of the persistent challenges faced by data aggregators has been managing disparate data sets for analysis and visualization. In the United States, the collection of public health data is governed at the local and state levels [50]. Strategies differ by state, with no central government authority to standardize data collection and reporting. The Council of State and Territorial Epidemiologists published standards for the clinical diagnosis of COVID-19 and data elements to report in April 2020, with updates in August 2020 and August 2021 [40,45,51]. The Council of State and Territorial Epidemiologists also recommended that states enact laws to make cases of COVID-19 reportable to public health authorities. The CDC has no authority to require reporting, stating “COVID-19 case surveillance data are collected by jurisdictions and reported voluntarily to CDC” [52].

Problems with data quality, standards, and availability have been described by dashboard and aggregator teams [53-56] and journalists [57-60]. Problems in data standardization and availability were somewhat alleviated during the first year of the pandemic, but data on case counts became unreliable by early 2022 because of the introduction of rapid at-home test kits [61,62].

Data that are visualized by a person or an organization that did not originally collect the data is an example of data reuse. The movement around Findable, Accessible, Interoperable, Reusable (FAIR) data includes the responsibility of providing appropriate data citations so that the original source and providence of data are discoverable [63]. The disconnect between the vision of the FAIR data and the findings of this survey is important. One challenge is that COVID-19 data are obtained in frequent updates (rather than from archived data sets) and often from data aggregators. This highlights the gap between the real-world need for trustworthy display of data in public health and typical use cases for using FAIR principles.

Aligning Visualization Goals and Visualized Data

What are the purposes of public-facing visualizations of pandemic data, and what data are needed to achieve those purposes?

Dashboards are often described as tools to support decision-making. Visualizations have played an important role in educating citizens about the pandemic and therefore may encourage changes in behavior to mitigate transmission. However, visualizations are likely to have a constellation of purposes. For example, a visualization could help establish trust between public health authorities and citizens. Further, effectively promoting behavior change may depend on first conveying the magnitude of human suffering caused by the pandemic.

The question of what data are useful for decision-making was addressed early in the pandemic by former CDC Director Dr Tom Frieden. He argued that there is a mismatch between the most commonly available data—counts of cases, hospitalizations, and deaths—and the data that are the most useful for guiding COVID-19 response in communities. He suggested that local decision-making for formulating policies should use data that include the number of unlinked infections, number of health care worker infections, and trends in excess mortality [64].

Visualizations as Arguments

Data visualizations are often assumed to be neutral and objective mechanisms of communication, but they are not. Designing and developing visualizations require numerous decisions regarding the selection of data and methods of presentation. It has been argued that all visualizations are rhetorical and therefore have the power to influence beliefs and behaviors [65,66].

In the context of the COVID-19 pandemic, public health authorities and government officials have made decisions about what data to collect and what data to not collect. These decisions constrain the messages that visualizations can send. In addition, the messages from these visualizations may imply a sense of authority and certainty through their association with organizations that have traditionally been respected (public health agencies, universities, and news organizations) and the “clean lines and structured layouts of traditional visualizations” [65]. This authority and certainty may obscure the extent of human suffering caused by COVID-19, echoing concerns raised by Dragga and Voss [67] in their analysis of graphs depicting fatalities and injuries from causes such as industrial workplaces and baby walkers.
In the United States, the authority of COVID-19 visualizations and the data behind them have been questioned, with various groups asserting that the severity of the pandemic has been either overplayed or downplayed. As many state and local policies for reopening schools and businesses were commonly tied to metrics about the pandemic, such as the test positivity rate or hospitalization rate, people tired of pandemic restrictions have accused COVID-19 data and dashboards of becoming political tools to prevent a return to normal. Other groups adopted a different perspective. For example, in April 2022, a coalition of public health practitioners, scientists, health care workers, educators, and advocates known as The People’s CDC released a statement criticizing the new definitions for categories of community transmission rates. They wrote the following:

The resulting shift from a red map to a green one reflected no real reduction in transmission risk. It was a resort to rhetoric: an effort to craft a success story that would explain away hundreds of thousands of preventable deaths and the continued threat the virus poses. [68]

The Connection Between Data, Usability, and Understandability

Public-facing visualizations of pandemic data are useful only if viewers are able to understand and interpret the data displays they see. Dashboard designers might choose to display large amounts of data with the goal of allowing the viewers to come to their own interpretation of the data without the prescriptive guidance of dashboard designers. However, this effort at transparency can backfire if the viewers are overwhelmed by the complexity or arrive at incorrect conclusions [25,65]. Viewers may assume that websites with more data are more accurate, but the volume of data and visualizations may obscure uncertainties in the data.

Visualization and communication researchers play crucial roles in determining how to better design public-facing dashboards for infectious disease data. Several studies have used COVID-19 data and dashboards in user studies [23-25,69]. Identifying best practices will accelerate the development of effective dashboards and trackers, and the software tools commonly used by public health authorities could incorporate those recommendations into templates. An important area for future investigation is determining if effective design practices for COVID-19 data can be applied to display other types of public health data.

Current research in the field of visualization seeks to develop software tools to assist nonexpert users in choosing effective visualization techniques to support their specific data sets and goals (as demonstrated in studies by Lavalle and Mate [70] and Golfarelli and Lizzi [71]). This aligns with 2 of the themes from this study, choosing the values to display and choosing the graphical form of the visualization. These studies are often based in the domain of business analytics; however, future work could focus on the domain of public health.

Limitations

This study was limited to dashboards and trackers available to the public as of August 2020 and therefore does not include dashboards used internally by health care and public health organizations. It excludes visualizations produced exclusively for smartphone apps and visualizations that focus on specific populations, such as nursing homes or prisons, or nontraditional data types, such as wastewater sampling.

Conclusions

This analysis reveals the extent to which dashboards and trackers informing the American public about the COVID-19 pandemic relied on an ad hoc pipeline of data sources and data aggregators. The pandemic has been characterized by disparate and evolving data standards, which has complicated the development of dashboards and trackers that display data over time and across regions. The 128 websites of dashboards and trackers identified in this survey offer an opportunity to compare different approaches to the display of similar data. This work highlights examples that provide clarity in interpreting data, and those that obscure the meaning of the data and may potentially mislead viewers.

Conflicts of Interest

None declared.

Multimedia Appendix 1

State-focused, nationwide, and global dashboards and trackers were examined. Includes the URL for each site, data sources, and type of visualization tool or method used. Government websites are listed in gray shading.

[PDF File (Adobe PDF File), 336 KB - humanfactors_v10i1e43819_app1.pdf]

Multimedia Appendix 2

State-focused, nationwide, and global dashboards and trackers for vaccination on websites that are not included in Multimedia Appendix 1. Includes the URL for each site, data sources, and type of visualization tool or method used. Government websites are listed with gray shading.

[PDF File (Adobe PDF File), 176 KB - humanfactors_v10i1e43819_app2.pdf]

Multimedia Appendix 3

Web pages for data sources and technical information provided by prominent data aggregators and dashboard developers.

https://humanfactors.jmir.org/2023/1/e43819
References


29. About Us. 1Point3Acres. URL: https://corona1point3acres.com/en/about [accessed 2023-01-02]


33. COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. GitHub. URL: https://github.com/CSSEGISandData/COVID-19 [accessed 2023-01-02]


64. Former CDC Director and Resolve to Save Lives president and CEO, Dr. Tom Frieden, warns of COVID-19 data myths and misuses, outlines metrics that matter. Prevent Epidemics. 2020 Jun 11. URL: https://tinyurl.com/2p8rb6xb [accessed 2023-01-02]


68. The People’s CDC. The CDC is beholden to corporations and lost our trust. We need to start our own. The Guardian. 2022 Apr 3. URL: https://www.theguardian.com/commentisfree/2022/apr/03/peoples-cdc-covid-guidelines [accessed 2023-01-02]


Abbreviations

API: application programming interface
CDC: Centers for Disease Control and Prevention
CSSE: Center for Systems Science and Engineering
FAIR: Findable, Accessible, Interoperable, Reusable
JHU: Johns Hopkins University
The Teach-ABI Professional Development Module for Educators About Pediatric Acquired Brain Injury: Mixed Method Usability Study

Lauren Saly¹, BA, BEd, OCT, MA; Christine Provvidenza¹,²*, MSc; Hiba Al-Hakeem¹, MA; Andrea Hickling¹,³, MScOT, OT Reg (Ont); Sara Stevens¹,², PhD, C Psych; Lisa Kakonge¹,⁴,⁵, MSc, Reg CASLPO, CCC- SL P; Anne W Hunt¹,⁶, PhD, OT Reg (Ont); Sheila Bennett⁷, EdD; Rhonda Martinussen⁸, PhD; Shannon E Scratch¹,²,⁶,⁹, PhD, C Psych

1Bloorview Research Institute, Holland Bloorview Kids Rehabilitation Hospital, Toronto, ON, Canada
2Holland Bloorview Kids Rehabilitation Hospital, Toronto, ON, Canada
3Department of Occupational Science and Occupational Therapy, Faculty of Medicine, University of Toronto, Toronto, ON, Canada
4School of Rehabilitation Science, McMaster University, Hamilton, ON, Canada
5Department of Speech-Language Pathology, University of Toronto, Toronto, ON, Canada
6Rehabilitation Sciences Institute, University of Toronto, Toronto, ON, Canada
7Department of Educational Studies, Brock University, St. Catherines, ON, Canada
8Department of Applied Psychology and Human Development, Ontario Institute for Studies in Education, University of Toronto, Toronto, ON, Canada
9Department of Paediatrics, University of Toronto, Toronto, ON, Canada
*these authors contributed equally

Corresponding Author:
Shannon E Scratch, PhD, C Psych
Bloorview Research Institute
Holland Bloorview Kids Rehabilitation Hospital
150 Kilgour Road
Toronto, ON, M4G 1R8
Canada
Phone: 1 416 425 6220 ext 3261
Email: sscratch@hollandbloorview.ca

Abstract

Background: Acquired brain injury (ABI) is a leading cause of death and disability in children and can lead to lasting cognitive, physical, and psychosocial outcomes that affect school performance. Students with an ABI experience challenges returning to school due in part to lack of educator support and ABI awareness. A lack of knowledge and training contribute to educators feeling unprepared to support students with ABI. Teach-ABI, an online professional development module, was created to enhance educators’ ABI knowledge and awareness to best support students. Using a case-based approach, Teach-ABI explains what an ABI is, identifies challenges for students with ABI in the classroom, discusses the importance of an individualized approach to supporting students with ABI, and describes how to support a student with an ABI in the classroom.

Objective: This study aims to assess the usability of and satisfaction with Teach-ABI by elementary school educators. The following questions were explored: (1) Can elementary school teachers use and navigate Teach-ABI?, (2) Are the content and features of Teach-ABI satisfactory?, and (3) What modifications are needed to improve Teach-ABI?

Methods: Elementary school educators currently employed or in training to be employed in Ontario elementary schools were recruited. Using Zoom, individual online meetings with a research team member were held, where educators actively reviewed Teach-ABI. Module usability was evaluated through qualitative analysis of think-aloud data and semistructured interviews, direct observation, user success rate during task completion, and the System Usability Scale (SUS) scores. The usability benchmark selected was 70% of participants performing more than half of module tasks independently.

Results: A total of 8 female educators participated in the study. Educators were classroom (n=7) and preservice (n=1) teachers from public (n=7) and private (n=1) school boards. In terms of task performance, more than 85% of participants (ie, 7/8) independently completed 10 out of 11 tasks and 100% of participants independently completed 7 out of 11 tasks, demonstrating...
achievement of the module usability goal. The average overall SUS score was 86.25, suggesting a high satisfaction level with the perceived usability of Teach-ABI. Overall, participants found Teach-ABI content valuable, useful, and aligned with the realities of their profession. Participants appreciated the visual design, organization, and varying use of education strategies within Teach-ABI. Opportunities for enhancement included broadening content case examples of students with ABI and enhancing the accessibility of the content.

**Conclusions:** Validated usability measures combined with qualitative methodology revealed educators’ high level of satisfaction with the design, content, and navigation of Teach-ABI. Educators engaged with the module as active participants in knowledge construction, as they reflected, questioned, and connected content to their experiences and knowledge. This study established strong usability and satisfaction with Teach-ABI and demonstrated the importance of usability testing in building online professional development modules.

*(JMIR Hum Factors 2023;10:e43129)* doi:10.2196/43129

**KEYWORDS**
acquired brain injury; educators; professional development; usability testing; satisfaction testing; knowledge translation; usability; death; disability; children; development; Ontario; research; online; school

**Introduction**

**Background**

Acquired brain injury (ABI) is defined as temporary or permanent damage to the brain that occurs after birth from a traumatic brain injury (TBI) or non-TBI [1]. ABI is a leading cause of death and disability in children [2]. After sustaining an ABI, outcomes vary based on several individual and injury-related factors, such as personality, preinjury strengths and needs, and location and severity of the injury [1-3]. Mild (ie, concussion), moderate, and severe brain injuries can lead to a variety of lasting cognitive, physical, and psychosocial outcomes [4-6] that affect students’ school performance [7,8].

Globally, educators report feeling underprepared and unaware of how to support students with ABI in the classroom [8-10]. School-aged children with ABI often experience challenges returning to school due in part to a lack of educator support and awareness of ABI [1]. Separate from an educators’ teaching approach, having ABI knowledge has been shown to influence academic and social domains for students [11]. Moreover, students with ABI report greater life satisfaction when their teachers are understanding of their needs and provide encouragement [11]. Therefore, a supportive school environment can facilitate successful school reintegration for children with ABI [12].

**The Canadian Context**

In Canada, children with ABI are a “silent voice” in the education system [13]. Outside of exploring additional education or training specific to developmental disorders, educators do not receive adequate instruction related to ABI during preservice training or as practicing professionals [14]. For example, in Ontario, Canada’s most populated province, the Education Act (1980) [15] separates students with special education needs into 5 broad categories: intellectual, behavioral, communication, physical, and multiple. ABI is not a separate category, and unfortunately, the evolving nature of ABI may make it difficult to fit students’ areas of need into a distinct category [1]. For example, the diverse range of ABI symptoms includes a combination of cognitive, physical, psychosocial, and communication concerns. Hence, ABI is a unique exceptionality due to its wide, significant, and individualized impact across many domains of functioning [14]. Identification within a category equips educators with additional knowledge and awareness of strategies to support students within the Ontario education system. In 2018, the passing of Bill 193, also known as Rowan’s Law [16], mandated requirements to enhance concussion safety in Ontario. The act was created to raise awareness about concussion and improve concussion safety within amateur competitive sport by mandating sport organizations to (1) have athletes review concussion awareness and education resources approved by the Minister of Tourism, Culture and Sport; (2) develop a concussion code of conduct and have athletes review the code; and (3) establish a removal-from-sport and a return-to-sport protocol [16]. Many Ontario school boards responded to Rowan’s Law by implementing yearly concussion training for educators; however, evidence suggests that this training is brief and focuses on signs and symptoms, rather than addressing potential long-term impacts and how to support deficits [17]. Furthermore, concussion is only 1 condition under the diverse umbrella of ABI. Therefore, a gap in training related to mild, moderate, and severe ABI remains. Recently, Stevens and colleagues [10] confirmed that Ontario educators lack the knowledge and confidence to support students with ABI in the classroom. Ontario educators also reported the need for a course to improve their knowledge and awareness of pediatric ABI. Researchers at the Holland Bloorview Kids Rehabilitation Hospital (HBKRH) in Toronto, Ontario, responded to this need by creating an online professional development module called Teach-ABI.

**Development of Teach-ABI**

**Overview**

The creation of Teach-ABI used an integrated knowledge translation (iKT) approach [18] and 2 process models throughout the design and testing phases: (1) Kern’s (2009) Six-Step Approach for Curriculum Development for Medical Education was used to develop Teach-ABI content and format [19]; and (2) the Knowledge-to-Action cycle [20] was used to consider the broader environment and context of this module. Importantly, Ontario educators were engaged as end users to
co-design Teach-ABI to maximize usability and relevance in the education setting. Applying these process models resulted in 6 phases of module development: (1) problem identification, needs assessment, and an environmental scan; (2) curriculum development (eg, content and delivery); (3) usability testing; (4) pilot testing; (5) efficacy testing and preimplementation planning; and (6) sustainability planning and generalizability. This paper summarizes phases 1 and 2 of Teach-ABI development, and discusses the methodology and findings from the usability testing of Teach-ABI (phase 3). Phases 4-6 are planned as future work.

**Phase 1: Problem Identification, Needs Assessment, and an Environmental Scan**

The problem was identified and examined through a needs assessment workshop conducted with Ontario educators [10]. Educators confirmed the knowledge gap related to pediatric ABI and identified the need for a standardized, accessible, engaging, and short e-learning program that would help raise awareness and knowledge about pediatric ABI and the unique needs of these students in the classroom. An online format that incorporated a blended-learning approach, using instructional methods including videos and a case study, was suggested by educators [10]. A detailed environmental scan of publicly available resources was then conducted, with no existing resources meeting the identified need [21]. With this in mind, an interdisciplinary stakeholder group was formed to advise on the development of Teach-ABI. This stakeholder group included clinicians (eg, neuropsychologists, occupational therapist, speech language pathologist), researchers, a knowledge translation specialist, academic faculty of teacher’s colleges, teachers, and families and youth with lived experience of ABI.

**Phase 2: Curriculum Development**

The design of Teach-ABI involved defining specific and measurable learning objectives and developing educational strategies. Bloom’s Taxonomy [22] was used to inform the learning objectives of Teach-ABI, which focused specifically on fostering the remembering and understanding of information by end users. The established learning objectives of Teach-ABI were to (1) define ABI; (2) identify challenges for students with ABI in the classroom; (3) discuss the importance of taking an individualized approach to supporting students with ABI; and (4) describe how to support a student with an ABI in the classroom. These learning objectives formed the basis of the module content, which was developed by a practicing classroom teacher with specialized knowledge in pediatric ABI (LS). A knowledge-translation specialist (CP), an e-learning specialist, a graphic designer, and a videographer were engaged to develop the format of the module. Teach-ABI was created across multiple stages of iterative design and development in consultation with the interdisciplinary stakeholder group (2018-2019).

Teach-ABI is a self-directed, online module that provides information to educators about ABI causes and outcomes, and strategies for supporting students with ABI in the classroom. Given the broad developmental needs of students, the first iteration of Teach-ABI is designed for elementary school educators in Ontario, Canada. Teach-ABI introduces the concept of ABI and provides examples of potential challenges after an injury and appropriate strategies to support students in the classroom. Teach-ABI uses a case study design with links to websites and resources, embedded videos, and downloadable information sheets. The case study follows the story of Olivia, a grade 4 student who sustained an ABI at age 5, and Mr. H, her teacher, who learns how to support Olivia over time. Teach-ABI is divided into 2 parts: (1) an overview of ABI and the presentation of a student with an ABI in the classroom; and (2) ways to support a student with an ABI by providing classroom strategies for cognitive, emotional, physical, behavioral, and communication outcomes. See Figure 1 for screenshots showcasing different components of Teach-ABI.
**Phase 3: Usability Testing**

The primary focus of this study was to engage Teach-ABI end users, elementary school educators, to determine the usability of and satisfaction with the Teach-ABI module. Usability was conceptualized as a user’s experience with Teach-ABI, guided by questions used in previous investigations of online learning products: “Does the e-learning [resource] function as designed and intended?”; “Can learners interact with and navigate around as they need to?” [23]. This study focused on the perceived usability (ie, ease of use and navigation of the interface) and satisfaction (ie, subjective experience of end users) with the module design and content, as these aspects can affect users’ comprehension and application of information [24,25].

Given this, there were 3 main research questions:

- Can participants use and navigate Teach-ABI?
- Are the Teach-ABI content and features satisfactory?
- What modifications are needed to improve Teach-ABI?

**Methods**

**Design and Participants**

A mixed method prospective study design was used. Elementary educators were recruited including preservice teachers, classroom teachers, special education teachers, principals, vice principals, registered early childhood educators, and educational assistants. The authors acknowledge that the word educator is an umbrella term that encompasses teachers and other education-related professionals, the same way that the school environment is a term that includes the classroom and other aspects of school such as the playground. For the remainder of this paper, we will be using these terms interchangeably, in a similar fashion as our participants, to best reflect participant data.

Individuals were eligible to participate in the study if they met the following criteria: (1) currently enrolled in a teacher’s college program that will provide certification as an elementary school teacher with the Ontario College of Teachers; or (2) currently registered with the Ontario College of Teachers as an elementary school teacher (primary/junior or junior/intermediate teaching qualifications); or (3) currently working in an Ontario public elementary school as an educational assistant or registered early childhood educator. None of the interested participants met the following exclusion criteria: (1) non-English speaking; or (2) had cognitive, physical, or visual impairments that would require accommodations to use Teach-ABI. Community sampling through research flyers, social media, and the HBKRH website was used.

**Ethics Approval**

Consent was obtained from all participants prior to commencing the study in compliance with the research ethics procedures (REB approval number 2020-0294-1588-2).
Data Collection and Outcomes

Each participant attended a virtual private meeting over Zoom (Zoom Video Communications, Inc.), an online videoconferencing platform that has been utilized and found to be effective for facilitating qualitative data collection [26]. All questionnaires were hosted on REDCap (Vanderbilt University), a web-based platform for creating and managing surveys and survey data [27]. The usability of the Teach-ABI module was evaluated through qualitative analysis of think-aloud data and semistructured interviews, direct observation, user success rate during task completion, and the System Usability Scale (SUS) [28]. The think-aloud method originated from cognitive interviewing and invites participants to verbally share what they are thinking, feeling, and doing as they complete a task [29]. Topic probes were adapted from past health information usability studies [30]. Participants were instructed to comment on the module’s content, images, features, ease of interface use, aspects they liked or disliked, and suggestions for improvement.

While participants reviewed the module, they shared their screen in the Zoom meeting with the research team member. This allowed the researcher to capture the usability of the module through direct observation and evaluating task completion, which have been used to investigate the usability of online resources [30,31]. Field notes were taken as participants navigated the module and referenced content, images, and features. Notes were also included when participant verbalizations were vague (eg, “I really like this part”) or when any areas of difficulty or confusion arose. These notes were combined with the think-aloud data to examine the research questions.

The researcher observed participants’ completion of 11 tasks related to Teach-ABI (see Table 1 for the task list) and rated their level of success based on 1 of 3 outcomes: completed with ease, completed with help, and did not complete.

Table 1. Teach-ABI\textsuperscript{a} task list.

<table>
<thead>
<tr>
<th>Task number</th>
<th>Task type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Access the module</td>
</tr>
<tr>
<td>2</td>
<td>Input information to create certificate</td>
</tr>
<tr>
<td>3</td>
<td>Browse content</td>
</tr>
<tr>
<td>4</td>
<td>Play the introduction video titled “Why the Teach-ABI Module Was Developed”</td>
</tr>
<tr>
<td>5</td>
<td>Download and open the tip sheet, titled “What Is Acquired Brain Injury”</td>
</tr>
<tr>
<td>6</td>
<td>Complete knowledge check (true or false)</td>
</tr>
<tr>
<td>7</td>
<td>Play the video titled “Supporting Students With ABI in the Classroom”</td>
</tr>
<tr>
<td>8</td>
<td>Hover over term to read definition of externalizing behaviors</td>
</tr>
<tr>
<td>9</td>
<td>Explore links outside the module and return back to the module</td>
</tr>
<tr>
<td>10</td>
<td>Navigating the module—return to previous slides</td>
</tr>
<tr>
<td>11</td>
<td>Access 1 or 2 resources in the resource list</td>
</tr>
</tbody>
</table>

\textsuperscript{a}ABI: acquired brain injury

The chosen tasks were characteristic of actions that must be completed to successfully engage with the module. Participants were instructed to navigate the module as they normally would, which involved minimal to no interference from the researcher. Thus, participants were not asked to complete the specific tasks, rather, the researcher observed their completion of the tasks without any direction.

After reviewing the module, participants completed the SUS and a semistructured exit interview. The SUS is a validated 10-item questionnaire that provides a quick assessment of a system or tool’s perceived usability [32-34]. While the questionnaire was modified to suit this study (see Multimedia Appendix 1), research has demonstrated that minor linguistic changes do not impact the validity of the scale [32]. A semistructured interview guide was used to learn more about participants’ experiences completing Teach-ABI. Topics were consistent with previous usability studies and guidelines [35-37] and included overall impression, liked and disliked aspects of the module, navigation experience and feature usability, and suggestions for improvement. The exit interview was audio-recorded.

Data Analysis

The average score on each of the 11 tasks was examined, in addition to participants’ individual scores on each task. A task was flagged as a usability problem if less than 70% of participants were able to complete it independently [30,38]. For this study, the usability goal was that more than half of the 11 tasks (ie, 6/11 tasks) would be completed independently by more than 70% (ie, n≥6/8) of participants. The ability to complete tasks needed to navigate Teach-ABI is a suitable way to determine module usability [39].

The SUS was scored using the steps outlined by Brooke [28]. Raw scores were converted to a total score out of 100. Scores were interpreted in relation to norm-referenced data, with an average score of 68 representing above average usability [33], and using a curved grading scale developed by Sauro and Lewis [40,41], which pairs scores out of 100 with a letter grade ranging from F (low) to A+ (high).
Two members of the research team (LS and HA-H) analyzed the qualitative data using content analysis [42,43]. Usability sessions and exit interviews were transcribed verbatim and reviewed multiple times for accuracy. Initial codes were generated from LS’s familiarization with the data and applied to the first transcript during line-by-line open coding. The initial list of codes was flexible and changed as the first transcript was coded. HA-H then coded the first transcript using the flexible list of codes. LS and HA-H discussed the existing codes and to create additional codes. These initial codes were used to code each transcript independently, and LS and HA-H met regularly to check for agreement related to the assigned codes and to create a final codebook. An explanation of each code was provided to ensure that they were applied consistently. Each code was also linked to one of the research questions to ensure the study focus remained central [44]. The codebook was flexible, as new codes were added throughout the coding process. Before a code was added, both researchers agreed on its inclusion and subsequent definition. The transcripts and codes were then organized in NVivo (QSR International), which was used to recode the transcripts based on the updated coding list. NVivo also provided structure and accessibility to the codes and meaning units within each code and allowed the data to be easily explored and reviewed to generate meaning and establish categories and subcategories [45]. The study reached data saturation as new or valuable information was not expected with additional interviews [46,47]. This is evident by the comprehensive information gathered when developing the categories and their relationships, as no new codes were identified following transcripts’ reviews.

**Results**

**Participant Demographics**

A total of 8 participants were enrolled in this study. The participants identified as female and were primarily early career practicing teachers employed by public school boards. Participants varied in their self-assessment of ABI knowledge. See Table 2 for participant information.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Sample (N=8), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender, female</td>
<td>8 (100)</td>
</tr>
<tr>
<td><strong>Profession</strong></td>
<td></td>
</tr>
<tr>
<td>Classroom teacher</td>
<td>7 (88)</td>
</tr>
<tr>
<td>Preservice teacher</td>
<td>1 (13)</td>
</tr>
<tr>
<td><strong>School employment setting</strong></td>
<td></td>
</tr>
<tr>
<td>Public</td>
<td>7 (88)</td>
</tr>
<tr>
<td>Catholic</td>
<td>1 (13)</td>
</tr>
<tr>
<td><strong>Years of professional experience</strong></td>
<td></td>
</tr>
<tr>
<td>0-5</td>
<td>6 (75)</td>
</tr>
<tr>
<td>6-10</td>
<td>0 (0)</td>
</tr>
<tr>
<td>11-15</td>
<td>1 (13)</td>
</tr>
<tr>
<td>16-20</td>
<td>0 (0)</td>
</tr>
<tr>
<td>21-25</td>
<td>1 (13)</td>
</tr>
<tr>
<td><strong>Experience with students with ABI</strong></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>4 (50)</td>
</tr>
<tr>
<td>No</td>
<td>4 (50)</td>
</tr>
<tr>
<td><strong>Prior experience completing an e-learning module</strong></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>4 (50)</td>
</tr>
<tr>
<td>No</td>
<td>4 (50)</td>
</tr>
<tr>
<td><strong>I feel that I have adequate knowledge about ABI</strong></td>
<td></td>
</tr>
<tr>
<td>Strongly disagreed/disagreed</td>
<td>4 (50)</td>
</tr>
<tr>
<td>Neutral</td>
<td>2 (25)</td>
</tr>
<tr>
<td>Strongly agreed/agreed</td>
<td>2 (25)</td>
</tr>
</tbody>
</table>

*ABI: acquired brain injury*
Task Performance

The average scores on each of the 11 tasks revealed that more than 85% (ie, 7/8) of participants independently completed 10 out of 11 tasks and 100% (8/8) of participants independently completed 7 out of 11 tasks. A usability problem occurred with downloading a tip sheet, with 5 participants able to independently download and open the tip sheet, 2 needing assistance, and 1 unable to complete this task. Overall, the study usability goal was met, as all participants completed more than 50% (>6/11) of tasks independently. See Table 3 for task performance scores.

Table 3. Task performance scores.

<table>
<thead>
<tr>
<th>Task</th>
<th>Completed with ease (N=8), n (%)</th>
<th>Completed with help (N=8), n (%)</th>
<th>Did not complete (N=8), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access the module</td>
<td>8 (100)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Input information to create certificate</td>
<td>7 (88)</td>
<td>0 (0)</td>
<td>1 (13)</td>
</tr>
<tr>
<td>Browse content</td>
<td>8 (100)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Play the introduction video titled “Why the Teach-ABI Module Was Developed”</td>
<td>7 (88)</td>
<td>1 (13)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Download and open the tip sheet, titled “What Is Acquired Brain Injury”</td>
<td>5 (63)</td>
<td>2 (25)</td>
<td>1 (13)</td>
</tr>
<tr>
<td>Complete knowledge check (true or false)</td>
<td>8 (100)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Play the video titled “Supporting Students With ABI in the Classroom”</td>
<td>8 (100)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Hover over term to read definition of externalizing behaviors</td>
<td>8 (100)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Explore links outside the module and return back to the module</td>
<td>8 (100)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Navigating the module—return to previous slides</td>
<td>8 (100)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Access 1 or 2 resources in the resource list</td>
<td>8 (100)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

aABI: acquired brain injury

SUS

The average overall score on the SUS was 86.25 (range 65-100), surpassing the above average score for system usability (ie, 68) [33]. Using the curved grading scale [34,40,41], 86.25 translated into a score of A+.

Qualitative Data

Overview

Qualitative content analysis yielded 5 categories and 13 subcategories (Figure 2).

Figure 2. Categories and sub-categories from interview data.
Content (Category 1): Value
Participants identified the module content as likable, understandable, and informative, and felt that the content was important to provide to educators. For example, one participant said, “Just talking about ABI in Ontario classrooms, I didn’t actually know that it’s not a recognized area of exceptionality and that it’s the leading cause of disability, so that is important for educators to know” [P6]. Similarly, participants shared that the content was informative, providing them with new knowledge: “This is new information to me. I just had no idea that there can be a delay in the challenges [after ABI]” [P2].

Content (Category 1): Relevance
As many as 7 participants felt the content was useful and aligned with the realities of their profession. For example, P2 stated “I found that it comes from a perspective where you understand the kids and you understand teachers and their perspective of the classroom.” Seven participants stated that they found the information to be useful, with 1 participant sharing their experience around downloading an information sheet, “Oh another sheet to download! I find these very useful” [P3].

Design (Category 2): Appearance
Participants found the module slides and tip sheets aesthetically pleasing (n=5) and enjoyed the colors (n=4) and pictures (n=4) used. For example, while navigating the module, participants shared: “I like the colour scheming so far” [P4] and “I like the use of pictures...It’s very visually appealing” [P6].

Design (Category 2): Organization
Participants appreciated that the module was divided into 2 parts: (1) ABI education and (2) supportive strategies (n=7). They also liked that the module had learning objectives and a content summary (n=5). They appreciated the use of bolding (n=5) and bullet points (n=6) to organize the slides, tip sheets, and videos. After reviewing the learning objectives, 1 participant shared: “I like the learning objectives. It really quantifies what am I going to get out of this module and makes it easier to make sure that I understand all of these steps by the time I’m done” [P5]. All participants appreciated the concision of the module components, expressing ideas such as, “I like that the videos are a reasonable amount of time” [P1]; “I like that [the information sheet is] short and easy to find the information” [P2]; “They’re to the point, easy to read, short, and won’t take up too much room on my computer” [P5].

Design (Category 2): Delivery
Teach-ABI presents information using various techniques, including interactive features (eg, reflection questions, knowledge check quizzes), videos, tip sheets, and the case study. Participants appreciated the varied techniques used and the engagement with the content that these techniques afforded. They expressed liking the case study approach to sharing information (n=6). When discussing the case study, participants stated: “The story of Olivia is great. It’s a nice way to follow something and to visualize it” [P4]; “I like the case study. It makes it more applicable and easier to understand” [P8]. Every participant shared their enjoyment of the knowledge check quizzes and a few highlighted that the quizzes made learning fun and accountable: “I did like the interactive pieces where you clicked to see the answer, or you dragged. Those things are fun!” [P4]; “I like the quizzes. They’re fun and they keep you accountable” [P7]. Participants enjoyed the lived experience videos (n=6) and described how the videos “humanize[d] the experience” [P5] and provided diverse perspectives of ABI (n=2).

Navigation (Category 3): Ease of Use and Access
The module was described as “user-friendly” (n=4), “easy to use,” (n=3), or “easy to follow” (n=3). Participants experienced little difficulty navigating the module and felt the features were simple to understand. They enjoyed the web-based nature of the module and its ease of use on personal devices: “I think a lot of modules I’ve used before open up in some weird flash player thing, so I liked that this was a web-based thing” [P8]. Participants found it easy to navigate between the slides and module sections (n=6), to access resource links that brought them outside the module, and to return to the module content: “I like that it’s hyperlinked so I can just easily access it” [P3].

Suggestions for Improvement (Category 4): Content
About one-third of the participants (n=3) suggested including additional examples of students with ABI to help broaden awareness of ABI and how it can affect children with different injuries; 2 participants shared a desire to learn more about the strategies listed to support students with ABI. Including resource links or videos that would provide further information about these strategies were suggested. For example, “I wish there was a resource that I could click there so I could learn more about that because that sounds interesting” [P8]. Participants expressed their enjoyment of the videos and the value they added to the module. Two participants frequently commented that it should be mandatory for participants to watch the videos with 1 sharing: “I think we should have to watch the videos” [P8].

Suggestions for Improvement (Category 4): Design
Two participants felt that the slides and text appeared small on their screen and suggested increasing font size, darkening the font, or creating a full-screen option that would expand the size of the slides. One participant noticed that the icon that invites users to download tip sheets was smaller than the other icons, making it seem less important. The same participant suggested making all of the icons the same size and adding an icon legend at the beginning of the module. Participants suggested reducing the amount of text on slides by using bullet points, charts, and images. One participant expressed, “A suggestion,...if [the slide] was bigger on the page, you could almost do a table with bullet points...often we deal with a lot of tables and bullet points...so it just becomes really easy to see” [P6]. Three participants mentioned adding a read-aloud feature, suggesting that it could improve users’ enjoyment of the module and meet educators’ different learning needs: “The one thing that I’ll say so far is that for some teachers, an audio feature, like listening to someone read it, would be really nice” [P6]. Participants commented on the length of time that it took to complete the module, sharing that most training completed at the beginning of the school year ranged from 15 to 30 minutes per module; yet the Teach-ABI module took longer to complete. They did
not believe the length of time to complete the Teach-ABI module was unreasonable but emphasized the timing inconsistency in comparison to other training modules. They suggested reducing text on slides and including a read-aloud function to lessen the time taken to complete the module.

**Suggestions for Improvement (Category 4): Navigation**

All participants commented on the simple navigation of the module, referring to the module as “user-friendly,” “easy to use,” or “easy to follow.” Three participants had trouble navigating a pop-up arrow and as described in the “Design” subcategory above, noted small suggestions to improve functionality (eg., increasing font size, adding read-aloud function, reducing text amount).

**Suggestions for Improvement (Category 4): Implementation Strategies**

Participants suggested a variety of implementation strategies, methods for enhancing the uptake or implementation of a program [48], that could facilitate use of the Teach-ABI module by educators. One participant suggested including the module within a paid professional development course related to special education. These optional courses, called Additional Qualifications in Ontario, provide specialized information to teachers, such as how to support students with special education needs. Another participant felt that the detailed information in the module would be very informative for preservice teachers learning how the Ontario education system works to support students with disabilities, and should be a mandatory component of training. Finally, 1 participant suggested using special education resource teachers (SERTs) to disseminate information about ABI to their staff. The SERTs would receive additional training about ABI and could share the module with colleagues during an in-service professional development day and address questions educators may have about ABI.

**Educator Reflections (Category 5): Content Engagement**

Participants commented on the module content and its consistency with their beliefs and knowledge, and made connections to their experiences in the classroom. They also discussed ways to apply the information moving forward. After reading the introduction to the case study, 1 participant shared, “Now I’m curious about what happened. It’s like you opened a book for me and there’s a story and if I don’t go on it’s like closing the book in the middle, so I want to go on and find out what he did” [P2]. Teachers also put themselves in the case study educator’s shoes and shared what they would do in this situation. They appreciated that the knowledge checks and reflection questions made them think and about the information: “I like that it’s got a question that makes you think, because if you think about it according to Mr. H’s approach, a lot of it is what most teachers would do” [P5]. They also reflected on the content and how it aligned with their professional knowledge and experiences (n=5).

For some participants, interacting with the module content led to realizations about their previous experiences in the classroom. Two participants shared that the module made them think that they may have taught students with ABI before, but they were not aware of this at the time. One participant stated, “It gets me thinking about some kids that I totally missed the boat on, thinking ‘oh I wish I had known this before’” [P1], while another shared “I have many students that have been in these situations that play sports and now I’m sitting here thinking how many of them could have had this as well” [P6].

Some participants who had previously taught students with ABI reflected on how the module information aligned with their personal experiences. For example: “In my past relationships with ABI, it’s been that situation where it’s misdiagnosed...teachers get confused and it is easier to just stamp them with something that gets them an IEP versus, identifying what is ABI” [P5]. This level of reflection was not noted from teachers without previous experience working with students with ABI.

**Educator Reflections (Category 5): Module Importance**

Participants reflected on the importance of the Teach-ABI module in relation to their lack of awareness and related training, and the contribution of Teach-ABI to their knowledge of ABI. Five participants discussed the lack of awareness about ABI that exists among educators. For example, 1 participant shared, “I know when I talk to other teachers, I hear false things all the time about concussions and ABI – well concussions – we don’t know anything about ABI. There’s definitely a lot of confusion about ABI in the classroom” [P2]. They agreed that educators are not provided with adequate training related to ABI. Many participants felt the module had a positive impact on their knowledge of ABI and believed it should be accessible to other educators. One participant stated, “It achieved its goal of educating teachers on what ABI looks like in the classroom and what responses were effective, while also being considerate to the fact that everyone’s experience is so individualized that it’s going to be something you learn as you go, but this is kind of a basis for what you can expect” [P5]. Another shared, “I think it’s something that would really help a lot of educators, like there’s not a lot of information about it. I learned a lot about ABI. I didn’t know, I would say, any of that. I definitely think there’s a lack of knowledge in education and I think teachers need to have access to [Teach-ABI] in some form or another” [P6].

**Discussion**

**Principal Findings**

This study describes the development of Teach-ABI and outcomes of usability testing. The primary aim of this work was to assess ease of use of the Teach-ABI interface, determine user satisfaction with the module design, and consider content modifications to optimize usability. The positive results expressed by the teacher participants regarding navigation of the module maintain that Teach-ABI is a highly usable, professional development resource.

Use and navigation of Teach-ABI were assessed through data triangulation across multiple sources: participant observation and task performance, think-aloud data, the SUS, and exit interview data. Results demonstrated usability and ease of navigation of the Teach-ABI module. Participants completed 7 out of the 11 selected tasks at a rate of 100%. The average score
on the SUS was 86.25, which is almost 20 points greater than scores awarded to systems with average usability [34]. Interview data also showed that participants described the module as easy to use and follow. Participant’s positive feedback on their experiences navigating the module contribute to an understanding of the module as user-friendly and highly learnable. Although a small number of participants encountered minor difficulties while navigating the module (eg, downloading the tip sheet, understanding pop-up arrow function), this can be remedied through minimal changes to the system.

In terms of Teach-ABI satisfaction, participants believed the module content was valuable, informative, and relevant to their profession. Module design, including the elements of appearance, organization, and delivery, was also rated positively. For example, participants appreciated the engaging features of the module (eg, case study, personal videos). Research highlights the importance of a module’s appearance and delivery, as it may influence users’ appeal and their ability to engage with the content [49]. Teacher participants also valued the self-reflection activities in Teach-ABI. Their reflections indicated that the module structure (eg, case study, knowledge checks) promoted a high level of user engagement and encouraged them to connect the topics to their practice.

The high level of satisfaction with the content and features of Teach-ABI speaks to the value of using an iKT approach combined with process models to inform product development [18-20]. Furthermore, the actions of reflecting on information, connecting it to previous experiences, and challenging existing knowledge link to higher-order thinking skills. This is significant, as eliciting higher-order thinking skills is associated with greater long-term recall and application of information [50]. In addition, educators were actively engaged with the module as they reflected on the content and made new connections to their personal teaching experience. Research has shown that reflection activities foster ongoing improvement in education practice and help to situate one’s context in their learning experience [51,52]. The inclusion of activities that end users can relate to through their teaching practice is a valuable design feature of online resources [51,53] and was the approach taken with the development of Teach-ABI.

**Considerations**

Access to online professional development opportunities, such as Teach-ABI, does not ensure that educators have the knowledge needed to support students with ABI in the classroom. However, it is important to acknowledge that teacher participants identified the potential for Teach-ABI to improve their knowledge and understanding of ABI. This finding is promising and consistent with previous research on ABI training and its association with increased educator knowledge, fewer ABI misconceptions, and higher levels of confidence related to teaching students with ABI [54,55]. Furthermore, preliminary research has examined the effect of an online TBI training module on educators’ knowledge and confidence related to supporting students with TBI. The results indicated that the online module significantly improved educators’ knowledge and confidence related to supporting students with TBI and this improvement was maintained at the 30-day follow-up [9,56].

Guided by an iKT approach [18], next steps involve understanding facilitators and barriers to implementing Teach-ABI; supports needed to foster implementation; and the impact of the module on shifts in knowledge, confidence, and teaching practices. In addition, considering how educators with different backgrounds (eg, previous work with students with ABI, familiarity with e-learning modules, and level of ABI knowledge) experience and benefit from Teach-ABI is an important future direction.

**Strengths**

The use of qualitative methods to examine usability was valuable, as it helped to explain the users’ response to content and features and situated their ratings of the module and suggestions for improving Teach-ABI. Besides, the sample size of the study is consistent with suggestions in the field of usability [57] and is suitable for reaching data saturation in qualitative interviewing [58]. Previous research studies examining training programs related to TBI have predominantly utilized quantitative methods, such as closed-ended surveys to examine usability [9,56], which provide a simple picture of usability. Qualitative methods helped achieve the program’s broader goal of creating a tool that is valuable and usable to Ontario educators and understanding participants’ experiences navigating Teach-ABI. In addition, the virtual data collection session resulted in an experience closely related to the real-world use of Teach-ABI by the study population. Instead of accessing Teach-ABI from a device provided by the researcher, participants accessed it using their own device and completed the module from a location of their choice, highlighting the ecological functionality of the module.

**Limitations**

There are some limitations related to the study sample. For example, all 8 participants were practicing (n=7) or preservice (n=1) classroom teachers. Originally, the study aimed to recruit Ontario educators broadly, including practicing and preservice classroom teachers, principals, educational assistants, and early childhood educators, to extend the generalizability of the results. All 8 participants identified as female. Although there are a significantly greater number of Ontario elementary educators that identify as female, as many as 4 times more female elementary school teachers than male teachers [59], the sample was not representative of the teacher population. Furthermore, usability research suggests that males evaluate e-learning systems differently than females [60]; therefore, future research should aim to capture male educators’ perspectives on Teach-ABI. In addition, the sample consisted of mostly early career classroom teachers. It would be important to capture the perspective of teachers with more than 5 years of experience to understand any differences in their experience completing and navigating virtual modules in comparison to educators in the beginning stages of their careers. Lastly, the sample characteristics were limited due to self-selection bias. Participation in the study was voluntary and participants were recruited through sharing the research flyer and information on the social media and website of a research hospital. Future research should target a broader group of educators using a wider variety of recruitment methods.
Conclusions
This study demonstrated strong usability and satisfaction with Teach-ABI, an innovative and novel online professional development module. Validated measures of usability combined with qualitative methodology revealed educators' high level of satisfaction with the design, content, and navigation of Teach-ABI. Educators engaged with the module as active participants in knowledge construction, as they reflected, questioned, and connected content to their experiences and knowledge. This study highlights the importance of usability testing in the build of online professional development modules. Furthermore, the comprehensive approach to testing the usability of Teach-ABI may be applied in future studies evaluating online modules.

Acknowledgments
Thank you to the teachers that participated in this study and made this research possible. We acknowledge the efforts of the members of the "Teach-ABI team," specifically Alicia Brown, Boey Ho, and Sarah Nauman. We also acknowledge the efforts of the members of the Neurorehab Outcomes via Education and Learning (NOvEL) Lab Team (Bloorview Research Institute), specifically Brendan Lam. This study was supported in part by funding from the Social Sciences and Humanities Research Council (SSHRC) and the Centre for Leadership at Holland Bloorview Kids Rehabilitation Hospital.

Authors' Contributions
LS and CP contributed equally to the manuscript. HA-H was the secondary data reviewer. HA-H and AH were involved in project and manuscript conception, drafted manuscript content, and approved the manuscript for final submission. SS, IK, AWH, SB, and RM were involved in project and manuscript conception, and drafted and approved manuscript content. SES is the senior author of the manuscript and as Lab Director, oversees all aspects of projects and student training.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Adapted SUS statements. SUS: System Usability Scale.

References
11. Wlodarczyk K. Educator evaluation of academic and social competence in students with acquired brain injury (ABI) relative to assessed performance and sense of belonging. Department of Graduate and Undergraduate Studies in Education, Faculty
29. Lewis C. Using the “thinking-aloud” method in cognitive interface design. Yorktown Heights, New York: IBM Thomas J. Watson Research Division; 1982. URL: https://domino.watson.ibm.com:8080/2513e49e05372ce852574ec0051eea4.html#~text=%22Thinking%20%5dloud%22%20is%20a,they%20work%20on%20a%20task] [accessed 2023-04-23]


40. Lewis JR, Sauro J. Can I Leave This One Out? The Effect of Dropping an Item From the SUS. Journal of Usability Studies 2017;13(1):38-46 [FREE Full text]


Abbreviations

ABI: acquired brain injury

HBKRH: Holland Blooview Kids Rehabilitation Hospital

iKT: integrated knowledge translation
**SERT:** special education resource teacher

**SUS:** System Usability Scale

**TBI:** traumatic brain injury
Acceptability and Potential Impact of the #chatsafe Suicide Postvention Response Among Young People Who Have Been Exposed to Suicide: Pilot Study

Background: Young people are more likely to be affected by suicide contagion, and there are concerns about the role social media plays in the development and maintenance of suicide clusters or in facilitating imitative suicidal behavior. However, social media also presents an opportunity to provide real-time and age-appropriate suicide prevention information, which could be an important component of suicide postvention activities.

Objective: This study aimed to test an intervention designed to equip young people to communicate safely online about suicide (#chatsafe) with a sample of young people who had recently been exposed to a suicide or suicide attempt, with a view to determining the role social media can play as part of a postvention response.

Methods: A sample of 266 young people from Australia, aged 16 to 25 years, were recruited to participate in the study. They were eligible if they had been exposed to a suicide or knew of a suicide attempt in the past 2 years. All participants received the #chatsafe intervention, which comprised 6 pieces of social media content that were sent to them weekly via direct message through Instagram, Facebook, or Snapchat. Participants were assessed on a range of outcome measures (social media use, willingness to intervene against suicide, internet self-efficacy, confidence, and safety when communicating about suicide on social media platforms) at baseline, immediately after the intervention, and at 4-week follow-up.

Results: After the 6-week #chatsafe intervention, participants reported substantial improvements in their willingness to intervene against suicide online, their internet self-efficacy, and their perceived confidence and safety when communicating about suicide online. Overall, the participants reported that it was appropriate to receive the #chatsafe intervention via social media, and no iatrogenic effects were recorded.

Conclusions: The findings suggest that it is safe and acceptable to disseminate suicide prevention information entirely via social media among young people who have recently been exposed to a suicide or suicide attempt. Interventions such as #chatsafe could potentially mitigate the risk of distress and future suicidal behavior in young people by improving the quality and safety of online communication about suicide and, as such, can be an important component of delivering a postvention response to young people.
Introduction

Background

Suicide is the leading cause of death among young people in Australia [1] and the second leading cause worldwide [2]. Although overall suicide rates have been decreasing in recent decades [3], this is not the case for young people for whom suicide rates have steadily increased in many parts of the world [4].

Youth suicides are between 2 and 4 times more likely to form part of a suicide cluster than adult suicides, with approximately 2.5% of youth suicides in Australia estimated to be part of a suicide cluster [5,6]. Suicide clusters are defined as a group of suicides that occur closer together in time and space than would normally be expected based on either statistical prediction or community expectation [7]. While the underlying mechanisms that facilitate the development and maintenance of suicide clusters are not well understood, one of the most common suggestions is that contagion or imitation occurs via social learning, where the suicide of one person may lead others who relate or identify with that person to engage in similar behavior [8,9]. Those thought to be most susceptible to this process are adolescents and young people [10] as well as those who are geographically close to the person who has died by suicide (e.g., witness the death), those who identify most closely with them, and those who are already susceptible in some way, (e.g., have a history of suicidality) [11].

One group who may be particularly susceptible to contagion are those who have been bereaved by, or exposed to, a suicide [12]. In a nationwide study conducted in Australia, almost 7% of young people aged 10 to 24 years who died by suicide had been exposed to the suicide of a friend or family member at some point in their lifetime [13], and exposure to a suicide has been shown to increase subsequent risk by approximately 300% [14]. Just as exposure can occur in person through connected networks, it can also occur via media (both traditional media and online media). Certain types of media reporting of suicide have been shown to increase imitative suicidal behavior in others [15], and being exposed to suicide in a way that glamorizes suicidal behavior or garners a lot of attention (e.g., public outpourings of how much someone will be missed) is thought to play a role in this [15].

Concerns relating to the impact of exposure to suicide have been heightened in the age of social media [9,16]. This is unsurprising, given the amount of time young people typically spend online and the speed at which unregulated and potentially distressing information about suicide can spread [17,18]. Concurrent with research findings for traditional media, exposure to graphic or distressing information about suicide on social media has been linked to an increase in suicidal thoughts and behaviors among young people [19]. This is worrying, given the rates at which young people are exposed to suicide-related content online, including graphic descriptions of suicide and statements encouraging someone to take their own life [20]. While some young people may actively seek suicide-related content online, in many cases, they are inadvertently exposed to this content [19,21,22].

Although exposure to suicide-related content online can be distressing, social media is also an important source of connection and support for young people, including when it comes to communicating about their own experiences with suicide [23,24] and grieving for someone who has died by suicide [25]. Therefore, social media is an important avenue to consider when supporting young people with their own suicidal thoughts and feelings as well as following bereavement by suicide. Indeed, social media platforms provide an opportunity to reach young people with suicide prevention information [20,26,27]; targeted information could be shared with those who have been bereaved by, or exposed to, suicide in an effort to provide support and minimize the spread of harmful or distressing information.

Very little is known about what constitutes the most effective postvention response for young people [28], and even less is known about how best to incorporate social media into those activities [9,29]. Although guidelines exist for implementing a multifaceted postvention response after a suicide has occurred [30-33], no postvention or cluster response strategy currently includes clear guidance for the use of social media. It has been argued that interventions that prevent the spread of harmful suicide-related content, particularly within 90 days of a suicide occurring, may have the potential to reduce the risk of subsequent suicide deaths within that community and provide necessary support to those exposed to the suicide [34]. Given its acceptability and its capacity to reach large numbers of young people quickly, social media could represent an important part of a postvention response.

One intervention that could form part of this response is #chatsafe. #chatsafe comprises a set of evidence-informed guidelines and accompanying social media campaign designed to educate young people about how to communicate safely online about suicide [26,27]. To date, the social media campaign has been viewed by more than 4 million young people worldwide [35]. It was evaluated among a general population sample of young people aged 16 to 25 years in Australia and was shown to increase participants’ perceived internet self-efficacy, confidence, and safety when communicating on social media about suicide. It also increased their willingness to intervene against suicide online [20]. However, to date, it has not been tested among young people who have previously been exposed to a suicide.

Aims and Hypotheses

The aim of this study was to test the #chatsafe intervention with a sample of young people who had been exposed to a suicide or suicide attempt in the past 2 years.
We hypothesized that, after receiving the #chatsafe intervention, young people who had been exposed to a suicide or suicide attempt in the past 2 years would report an increase in their willingness to intervene against suicide online (hypothesis 1). We also hypothesized that increases would be observed in participants’ perceived internet self-efficacy (hypothesis 2) as well as a greater adherence to communication behavior recommended by the #chatsafe guidelines (hypothesis 3). A further exploratory aim of this study was to investigate the safety and acceptability of the intervention and to determine whether age, gender, or rate of social media use influenced the impact of the #chatsafe intervention.

**Methods**

**Design and Setting**

This study largely used the same design as the original #chatsafe study [20], except that it sought to specifically recruit young people who had been exposed to a suicide or a suicide attempt (as opposed to the general population of young people). It used a prepost study design with a 6-week intervention period. The study was conducted online, and young people were assessed on the primary and secondary outcome variables at 3 time points: baseline (time 1; T1), immediately after the intervention (time 2; T2), and at the 4-week follow-up (time 3; T3). The participants also completed a short weekly survey, from week 1 to week 6. The study timeline is shown in Figure 1.

This study was conducted in Australia between July 2020 and March 2021. It has been reported in accordance with the Template for Intervention Description and Replication (TIDieR) checklist [36].

![Timeline of the study and #chatsafe intervention. W: week.](https://humanfactors.jmir.org/2023/1/e44535)

**Participants**

Young people were recruited to the study via targeted advertising on Instagram, Snapchat, and Facebook during the 5-month period from July to November 2020. Young people were eligible to participate if they (1) were aged between 16 and 25 years, inclusive; (2) lived in Australia; (3) had not participated in the previous #chatsafe study; (4) knew of someone who had died by suicide or attempted suicide in the past 2 years (including a friend, family member, or someone in their online or offline communities); and (5) were willing to provide the details of an active Instagram, Snapchat, or Facebook account to the research team to receive the intervention.

After providing consent, all communication with participants took place via direct message through their nominated social media platform. Young people were reimbursed Aus $30 (US $20.13) per completed survey via direct bank transfer.

**Intervention**

As described previously, the #chatsafe intervention comprises a set of evidence-informed guidelines that are distributed to young people via a co-designed suite of social media content [20,26,27]. For this study, 2 co-designed workshops were conducted in 2020 to create specific content for young people who had been impacted by a suicide or suicide attempt.

The intervention consisted of a 6-week social media campaign that was shared on the #chatsafe Instagram page [37]. Each week, 3 posts were shared on Instagram, resulting in 18 pieces of content in total. Not only were participants able to view the entire campaign on the public Instagram page but they were also sent 1 post per week via direct message to their preferred social media platform: Instagram, Facebook, or Snapchat. Information about available national support services and a link to a weekly acceptability questionnaire were also sent to participants each week. The intervention is described in Table 1, and specific examples of the content are shown in Figure 2.
Table 1. Content theme, content type, and information contained within the content and content copy of the intervention material.

<table>
<thead>
<tr>
<th>Week</th>
<th>Content theme</th>
<th>Content type</th>
<th>Information contained in content and content copy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>General introduction to the #chatsafe campaign</td>
<td>Text only</td>
<td>Introducing participants to the #chatsafe guidelines and how the content was developed</td>
</tr>
<tr>
<td>2</td>
<td>Safely sharing information about suicide: using trigger and content warnings</td>
<td>Text with digital illustration</td>
<td>Highlighting the importance of using trigger warnings with examples of how to do so</td>
</tr>
<tr>
<td>3</td>
<td>Self-care: take a break from social media</td>
<td>Digital illustration</td>
<td>Encouraging participants to take a break from social media after being exposed to upsetting content online</td>
</tr>
<tr>
<td>4</td>
<td>Language matters: how to safely talk about suicide online</td>
<td>Text only</td>
<td>Describing the importance of safe language when talking about suicide with examples of how to do so</td>
</tr>
<tr>
<td>5</td>
<td>Self-care: take a break from social media</td>
<td>Boomerang (no audio)</td>
<td>Encouraging participants to take a break from social media after being exposed to upsetting content online</td>
</tr>
<tr>
<td>6</td>
<td>How to check in on a friend affected by suicide</td>
<td>Animation</td>
<td>Normalizing the difficulty of talking about suicide and providing examples of how to check in on someone who has been affected by suicide</td>
</tr>
</tbody>
</table>

Figure 2. Examples of social media content shared on the #chatsafe social media pages during this study. Left: a text tile encouraging users to consider using a content warning. Middle: a still image of a short video (with no audio) depicting 2 young people “taking a break.” Right: a still image of an animation video discussing how to support someone affected by suicide.

Study Outcomes and Measures

The primary outcome of interest was participants’ willingness to intervene against suicide at T2, with the 2 subscales from this measure being perceived behavioral control and intent to intervene against suicide [38]. Secondary outcomes included internet self-efficacy [39] and perceived confidence and safety when communicating online about suicide [40] at T2. The measures used to assess these outcomes have been used previously and are described in the study by La Sala et al [20]. In brief, internet self-efficacy comprises 5 domains: reactive and generative (problem-solving and contributing unique information online), organization (organizing information on social media platforms), differentiation (willingness to follow hyperlinks in goal-oriented tasks), search (using advanced search engines), and communication (navigating social networking sites). Adherence to communication behaviors recommended in the #chatsafe guidelines was measured using items from the perceived safety questionnaire (eg, monitoring social media posts and reporting unsafe content) [26].

All data were collected through online self-report surveys at 3 time points using Qualtrics (Figure 1).

At T1, participants also completed a demographic questionnaire assessing age, primary language spoken at home, Aboriginal or Torres Strait Islander identity, gender identity, sexual orientation, student or employment status, and social media use [41].

Acceptability and safety of the #chatsafe intervention were also examined. Acceptability was assessed in 2 ways. First, participants were asked each week to complete a 5-point Likert emoji scale rating their satisfaction with the content sent to them that week [20]. Second, 5 purpose-designed questions assessing the overall acceptability of the 6-week intervention were included in the T2 survey. Safety was measured by the number and nature of serious adverse events and reactions to the content shared by the study team throughout the #chatsafe intervention.
Data Analysis
To test the primary hypothesis that there would be an increase in scores on both subscales of the willingness to intervene against suicide measure between T1 to T2, regression analyses were used to determine the extent to which the predictor variables (gender, age group, and social media use) could predict the primary outcome relative to no change. The changes in scores from T1 to T2 were grouped based on the magnitude of change from the baseline score, calculated from the SD of the baseline score multiplied by 0.3 (small to medium effect size as per Cohen classification [42]) to derive thresholds for substantial deterioration, no change, and substantial improvement (Multimedia Appendix 1). This standardized difference approach to effect size classification has been used in previous studies [43,44] and was also used to assess changes from T1 to T2 for the Internet Self-Efficacy Scale domains as well as changes from T1 to T3 for both the Willingness to Intervene Against Suicide and Internet Self-Efficacy measures. The thresholds used for these measurements are listed in Table S1 in Multimedia Appendix 1.

To assess the differences in both the primary and secondary outcome variables based on preidentified subgroups, the following subgroups were generated: gender (divided into male, female, and transgender and gender-diverse people), age group (younger participants aged 16-20 years and older participants aged 21-25 years), and time spent on social media (moderate social media users who spent <5 hours on social media per day and high social media users who spent more than 5 hours on social media per day).

Perceived safety, conceptualized as adherence to the #chatsafe guidelines, was calculated using items from the Perceived Safety Questionnaire at T2 and reported as frequencies and percentages, with Fisher exact test values reported where comparisons between T1 and T2 have been made. Evaluations of the #chatsafe intervention content at T2 were reported as frequencies and percentages.

Statistical analyses were conducted using StataIC 15 (StatCorp LLC) [45].

Ethics Approval and Safety
This study was approved by the University of Melbourne Human Research and Ethics Committee (ID: 1954623). In addition, several measures were taken to ensure participant safety. This included the development of an independent Safety Monitoring Committee to oversee study conduct, daily monitoring of all the #chatsafe social media accounts for any messages or comments that indicated distress, and monitoring of the weekly survey responses. Any distress reported by participants through contact with the study team or via responses to the weekly surveys was to be followed up within 24 hours. The participants were reminded that they were free to withdraw at any point and were also given the option of snoozing the weekly content, and this allowed them to take a 1-week break from the intervention.

All correspondence to the participants included contact details of age-appropriate support services, such as eheadspace and Kids Helpline.

Finally, adverse events (AEs) and serious adverse events (SAEs) were monitored. In accordance with the organization’s policies, AEs were defined as any untoward or adverse effect related or unrelated to the study (eg, comments that expressed suicidal ideation). SAEs were defined as an event that resulted in death or as immediately life threatening or required hospitalization [46].

Results

Demographic Details
As shown in Figure 3, a total of 1763 young people responded to the study advertisement and commenced eligibility screening; 454 young people were eligible and completed the T1 survey. Only participants who commenced the intervention and completed at least T1 and T2 were included in the analysis. This resulted in a final sample size of 266 and a retention rate of 58.59% across the study period.
The participant demographics are presented in Table 2. The participants were young adults aged between 16 and 25 years, with a median age of 18.9 years. Most of them (206/266, 77.4%) identified as cisgender female. More than half (145/266, 54.5%) of the sample identified as nonheterosexual, and the majority (213/266, 80.1%) were currently studying. Participants who did not complete the study and whose data were not retained in the final analysis did not significantly differ by age ($P=.62$), gender ($P=.90$), sexual orientation ($P=.12$), language ($P=.55$), Aboriginal and/or Torres Strait Islander descent ($P=.95$), student status ($P=.64$), relationship to someone who has attempted or died by suicide ($P=.85$), or social media use ($P=.19$).
Table 2. Demographic and baseline characteristics of participants who completed T1 and T2 (N=266).

<table>
<thead>
<tr>
<th>Baseline characteristics</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>18.9 (2.6)</td>
</tr>
<tr>
<td>Gender, n (%)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>32 (12)</td>
</tr>
<tr>
<td>Female</td>
<td>206 (77.4)</td>
</tr>
<tr>
<td>Transgender and gender-diverse participants</td>
<td>28 (10.5)</td>
</tr>
<tr>
<td>Sexual orientation, n (%)</td>
<td></td>
</tr>
<tr>
<td>Heterosexual (straight)</td>
<td>121 (45.5)</td>
</tr>
<tr>
<td>Lesbian or gay</td>
<td>14 (5.3)</td>
</tr>
<tr>
<td>Bisexual</td>
<td>74 (27.8)</td>
</tr>
<tr>
<td>Other</td>
<td>57 (21.4)</td>
</tr>
<tr>
<td>Language, n (%)</td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>240 (90.2)</td>
</tr>
<tr>
<td>Other</td>
<td>26 (9.8)</td>
</tr>
<tr>
<td>Aboriginal and/or Torres Strait Islander, n (%)</td>
<td></td>
</tr>
<tr>
<td>Aboriginal</td>
<td>5 (1.9)</td>
</tr>
<tr>
<td>Neither aboriginal nor Torres Strait Islander</td>
<td>261 (98.1)</td>
</tr>
<tr>
<td>Currently studying, n (%)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>213 (80.1)</td>
</tr>
<tr>
<td>No</td>
<td>53 (19.9)</td>
</tr>
<tr>
<td>Relationship to someone who has attempted or died by suicide, n (%)</td>
<td></td>
</tr>
<tr>
<td>Know in real life</td>
<td>234 (88)</td>
</tr>
<tr>
<td>Know via the internet</td>
<td>32 (12)</td>
</tr>
<tr>
<td>Social media use (hours), n (%)</td>
<td></td>
</tr>
<tr>
<td>&lt;1</td>
<td>4 (1.5)</td>
</tr>
<tr>
<td>1-2</td>
<td>42 (15.8)</td>
</tr>
<tr>
<td>2-3</td>
<td>80 (30.1)</td>
</tr>
<tr>
<td>3-4</td>
<td>74 (27.8)</td>
</tr>
<tr>
<td>≥5</td>
<td>66 (24.8)</td>
</tr>
</tbody>
</table>

The eligibility criteria meant that all participants had been exposed to a suicide or suicide attempt in the past 2 years. Most participants knew the person who had died by suicide or made a suicide attempt in their offline lives (234/266, 88%) as opposed to only knowing the person online.

Social Media Use
Social media use among the participants was high. More than half (154/266, 57.9%) of the participants reported that they spent 2 to 4 hours per day on social media, and almost one-fourth (66/266, 24.8%) reported spending >5 hours per day on social media. The most commonly used platform was Instagram, followed by Snapchat, YouTube, Facebook, and Twitter. Tumblr was the least-used platform.

Exposure to suicide-related content on social media was common (Table 3).
Table 3. Types of suicide-related content seen by young people in the previous 4 weeks at each time point (N=266 at T1 and T2, N=212 at T3).

<table>
<thead>
<tr>
<th>Types of Content</th>
<th>T1, n (%)</th>
<th>T2, n (%)</th>
<th>T3, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graphic descriptions of suicide</td>
<td>78 (29.3)</td>
<td>47 (17.7)</td>
<td>32 (15.1)</td>
</tr>
<tr>
<td>Graphic images of suicide</td>
<td>60 (22.6)</td>
<td>34 (12.8)</td>
<td>13 (6.1)</td>
</tr>
<tr>
<td>Means or methods of suicide</td>
<td>84 (31.6)</td>
<td>66 (24.8)</td>
<td>41 (19.3)</td>
</tr>
<tr>
<td>Plans of suicide</td>
<td>67 (25.2)</td>
<td>50 (18.8)</td>
<td>30 (14.2)</td>
</tr>
<tr>
<td>Statements that encourage people to take their own life</td>
<td>63 (23.7)</td>
<td>46 (17.3)</td>
<td>34 (16.0)</td>
</tr>
<tr>
<td>Statements that appear to deliberately seek to trigger difficult or distressing emotions in other people</td>
<td>108 (40.6)</td>
<td>74 (27.8)</td>
<td>60 (28.3)</td>
</tr>
<tr>
<td>Statements that include suicide pacts or suicide partners</td>
<td>27 (10.2)</td>
<td>17 (6.4)</td>
<td>16 (7.6)</td>
</tr>
<tr>
<td>Statements that place blame or make others feel responsible for another person’s safety</td>
<td>88 (33.1)</td>
<td>54 (20.3)</td>
<td>41 (19.3)</td>
</tr>
<tr>
<td>Statements that provide vulnerable people information about how to end their life</td>
<td>44 (16.5)</td>
<td>29 (10.9)</td>
<td>23 (10.9)</td>
</tr>
<tr>
<td>Suicide notes or goodbye notes</td>
<td>68 (25.6)</td>
<td>45 (16.9)</td>
<td>28 (13.2)</td>
</tr>
<tr>
<td>None</td>
<td>75 (28.2)</td>
<td>117 (44.0)</td>
<td>94 (44.3)</td>
</tr>
</tbody>
</table>

**Primary Outcome: Willingness to Intervene Against Suicide From T1 to T2**

Table 4 presents the results of the logistic regression analysis that examined predictors of improvement and deterioration, relative to no change, in both subscales of the Willingness to Intervene Against Suicide measure from T1 to T2.

Most (154/266, 57.9%) participants showed substantial improvement in perceived behavioral control, almost one-fifth (50/266, 18.8%) showed deterioration, and almost one-quarter (62/266, 23.3%) showed no change. Baseline perceived behavioral control was associated with significant improvement from T1 to T2, whereby higher baseline scores reduced the likelihood of significant improvement (odds ratio [OR] 0.92, 95% CI 0.89-0.96; \( P<.001 \)). No other predictor variables were associated with improvement in perceived behavioral control, and no predictor variables were associated with deterioration from T1 to T2.

Many (114/266, 42.9%) participants demonstrated improvement in intent to intervene, compared with 29.7% (79/266) of participants with no change in scores and 27.4% (73/266) who demonstrated deterioration. Of the potential predictors of improvement, only baseline intent to intervene scores were found to be significant, with higher baseline scores associated with a decrease in the likelihood of improvement (OR 0.90, 95% CI 0.87-0.95; \( P<.001 \)). No other variables were associated with improvement in intent to intervene, and no variables were associated with deterioration from T1 to T2.
### Table 4. Predictors of improvement and deterioration in the Willingness to Intervene Against Suicide (WIAS)-Perceived Behavioral Control (PBC) and Willingness to Intervene Against Suicide-Intent to Intervene T1 to T2.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Improvement&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Deterioration&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WIAS-PBC&lt;sup&gt;b&lt;/sup&gt;</td>
<td>WIAS-Int&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>OR (95% CI)</td>
<td>P value</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;21</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>≥21</td>
<td>0.99 (0.52-1.88)</td>
<td>.99</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Female</td>
<td>2.25 (0.88-5.71)</td>
<td>.09</td>
</tr>
<tr>
<td>Transgender and gender-diverse participants</td>
<td>1.63 (0.51-5.18)</td>
<td>.41</td>
</tr>
<tr>
<td>Sexual orientation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heterosexual or straight</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Lesbian or gay</td>
<td>4.21 (0.52-34.21)</td>
<td>.18</td>
</tr>
<tr>
<td>Bisexual</td>
<td>0.87 (0.43-1.74)</td>
<td>.69</td>
</tr>
<tr>
<td>Other</td>
<td>0.74 (0.34-1.61)</td>
<td>.45</td>
</tr>
<tr>
<td>Social media use (hours)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;5</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>≥5</td>
<td>1.03 (0.52-2.04)</td>
<td>.94</td>
</tr>
<tr>
<td>Baseline WIAS-PBC</td>
<td>0.93 (0.91-0.96)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Baseline WIAS-Int</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

<sup>a</sup>For both outcomes (improvement and deterioration), the comparator group consisted of participants who did not show a change in score over this period.

<sup>b</sup>WIAS-PBC: Willingness to Intervene Against Suicide–Perceived Behavioral Control.

<sup>c</sup>WIAS-Int: Willingness to Intervene Against Suicide–Intent to Intervene.

<sup>d</sup>OR: odds ratio.

<sup>e</sup>Row represents the reference group for the corresponding variable.

<sup>f</sup>Italicized values indicate significance.

### Secondary Outcomes

#### Willingness to Intervene From T1 to T3

Secondary analyses examining change in perceived behavioral control from T1 to T3 similarly found substantial improvement in most participants (139/212 65.57%); fewer than one-fifth demonstrated no change (35/212, 16.51%) or deterioration (38/212, 17.92%). Table S3 in Multimedia Appendix 1 shows the predictors of improvement in the perceived behavioral control subscale of the Willingness to Intervene Against Suicide measure.

A secondary analysis of the change from T1 to T3 indicated that half (104/212, 49.06%) of the sample were more likely to intervene, whereas approximately one-fourth demonstrated either no change (58/212, 27.36%) or deterioration (50/212, 23.58%). The predictors are presented in Table S3 in Multimedia Appendix 1.

### Internet Self-efficacy

Approximately one-third of the participants demonstrated improved reactive self-efficacy (85/266, 32.2%), differentiation self-efficacy (79/266, 29.7%), and organizational self-efficacy (81/266, 30.45%), and approximately one-fifth demonstrated improvement in communication self-efficacy (55/266, 20.75%) and search self-efficacy (51/266, 19.25%). Most participants demonstrated no change in subdomains of the Internet Self-Efficacy scale. The predictors of improvement and deterioration are listed in Table S4 in Multimedia Appendix 1. Higher baseline scores in each of the subdomains were associated with a reduced likelihood of improvement for the corresponding subdomain, whereas higher baseline scores in the differentiation and search subdomains were associated with deterioration in the differentiation and search domains, respectively (Table S5 in Multimedia Appendix 1). Being aged ≥21 years was also associated with a reduced likelihood of deterioration by 53% (OR 0.43, 95% CI 0.20-0.93; P=.03) in the reactive subdomain.
Confidence and Safety (Adherence to the #chatsafe Guidelines) When Communicating Online About Suicide

At each time point, the participants were asked about their online experiences and behaviors in the preceding 4 weeks. Almost two-thirds of the sample reported that they had liked, shared, or created suicide-related content at T1 (173/266, 65.04%) and at T2 (179/266, 67.29%). Of these participants, the proportion that indicated that they monitored their posts for unsafe content increased from T1 (113/173, 65.32%) to T2 (149/179, 83.2%). Many participants reported not seeing unsafe content on their posts at both time points (T1: 46/113, 40.71% and T2: 67/149, 44.08%).

Only those who reported seeing unsafe content were asked how they dealt with that content. Participants most commonly reached out to the person who posted across both time points, although the proportion decreased from T1 to T2 (T1: 38/113, 33.63%; T2: 38/152, 25%). Participants also reported that they deleted (T1: 32/113, 28.32%; T2: 42/152, 27.63%) or hid the post (T1: 26/113, 23.01%; T2: 25/152, 16.45%). Some participants also endorsed informing a trusted friend or adult (T1: 44/221, 19.91%; T2: 47/195, 24.10%) or contacting the relevant platform safety center (T1: 39/221, 17.65%; T2: 40/195, 20.51%), and a minority reported seeking professional advice (T1: 12/221, 5.43%; T2: 22/195, 11.28%). At each time point, most participants indicated that they thought about whether they felt able to respond to the individual before deciding whether to respond (T1: 147/221, 66.52%; T2: 145/195, 74.36%).

Acceptability of the #chatsafe Intervention

Weekly Acceptability of Intervention Content

Overall, participants responded positively to the intervention content sent each week, and at no point was the intervention content deemed unsafe. Participants responded most positively to content from week 6, “How to check in on a friend who has been affected by suicide,” and responded least positively to content from week 5, “self-care.” Acceptability did not vary by gender, age group, or level of social media use (Table 5).

Table 5. Weekly acceptability of #chatsafe intervention content.

<table>
<thead>
<tr>
<th>Week</th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positive, n (%)</td>
<td>Negative, n (%)</td>
<td>Positive, n (%)</td>
<td>Negative, n (%)</td>
</tr>
<tr>
<td>1</td>
<td>201 (90.95)</td>
<td>8 (3.62)</td>
<td>152 (68.77)</td>
<td>37 (16.74)</td>
</tr>
<tr>
<td>2</td>
<td>137 (95.81)</td>
<td>3 (2.1)</td>
<td>105 (73.43)</td>
<td>24 (16.79)</td>
</tr>
<tr>
<td>3</td>
<td>107 (83.59)</td>
<td>9 (7.03)</td>
<td>82 (64.07)</td>
<td>29 (22.65)</td>
</tr>
<tr>
<td>4</td>
<td>113 (87.60)</td>
<td>8 (6.21)</td>
<td>96 (74.42)</td>
<td>20 (15.51)</td>
</tr>
<tr>
<td>5</td>
<td>86 (69.92)</td>
<td>15 (12.19)</td>
<td>69 (54.09)</td>
<td>40 (32.53)</td>
</tr>
<tr>
<td>6</td>
<td>119 (96.75)</td>
<td>2 (1.62)</td>
<td>100 (81.30)</td>
<td>14 (11.38)</td>
</tr>
</tbody>
</table>

aWhat did you think about the campaign content this week?
bWould you share this week’s campaign content with your contacts on social media?
cHow did the campaign content you received today make you feel?
dTotal number of responses received in that week.

Table 5. Weekly acceptability of #chatsafe intervention content.

Postintervention Acceptability

Almost half (132/266, 49.62%) of the participants reported finding the #chatsafe content to be helpful. Almost half (126/266, 47.37%) of the participants reported that the intervention material made them feel more confident when talking about suicide online. Most participants reported that the #chatsafe content posed no risk to themselves (254/266, 95.49%), and they did not feel that it would be a risk to others (224/266, 84.21%). More than one-third (106/266, 39.85%) of participants believed that the #chatsafe content would help prevent further suicide or suicide attempts in others following an index suicide in the community.

Safety of the #chatsafe Intervention

No AEs or SAEs were observed during the study period. A total of 32 people were lost to follow up throughout the study period (ie, they changed their social media handle, deactivated their...
social media account, or unfollowed the #chatsafe profile and therefore could not be contacted). Across the 6-week intervention, 3 participants requested to "snooze" the content for a period of 1 week. None of the participants expressed distress or requested that a member of the study team contact them at any stage of the study.

Discussion

Principal Findings

The aim of this study was to explore the role social media can play in supporting young people who have been exposed to a suicide or a suicide attempt by testing the impact of the #chatsafe intervention. The findings from this study not only support the safety, acceptability, and impact of the #chatsafe intervention but also point to an increase in participants’ willingness to intervene against suicide online. The findings suggest that the #chatsafe intervention may have increased some young people’s internet self-efficacy as well as their confidence and safety when communicating online about suicide. Although most participants reported improvements in the primary and secondary outcome variables, they appeared to be quite proficient in safe communication practices at baseline, with high scores on perceived internet self-efficacy and a strong endorsement of items from the #chatsafe guidelines [26]. Although this limited the rate at which improvement on these outcomes could be measured, the findings from this study support the utility of using social media to reach young people with suicide prevention information.

Young people are frequently exposed to suicide-related content online, and it is well documented that exposure can increase the risk of future suicide and suicide-related behavior [8,9]. Almost two-thirds of the participants in this study had liked, shared, or created suicide-related content on social media, and the majority had seen posts online which made them think someone was at risk of suicide. High rates of exposure to content such as information about methods of suicide, statements that participants felt were deliberately attempting to trigger difficult or distressing emotions, and statements that made others feel responsible for someone’s safety were also recorded. Approximately one-fourth of the sample had seen suicide notes, comments encouraging suicide, and graphic images of suicide. This is of concern, considering that harmful content, such as specific details about suicide, is thought to encourage imitative behavior [14,34,47]. These data speak to the amount and type of suicide-related content that young people are exposed to online and add further weight to the growing concerns about the potential impact of social media on youth mental health and suicide risk [9,16,47]. Taken together, these data highlight the importance of equipping young people with the skills to keep themselves and their peers safe when actively or passively engaging with suicide-related information on social media. They also support social media being an important context to consider when implementing an effective postvention response for young people [20,32,48].

Implications

Findings from this study suggest that the #chatsafe intervention achieved its objective of educating young people about the importance of safe online communication about suicide and provides an example of how social media content could be incorporated into a postvention approach. The greatest increases were observed in participants’ perceived behavioral control to respond to suicide-related content online, suggesting that the #chatsafe intervention increased their belief in their ability to safely manage or intervene against suicide-related content. Equipping young people with the knowledge to keep themselves and others safe is the primary goal of #chatsafe and ensuring that young people feel able to share and respond to suicide-related content safely is the first step. However, despite most participants reporting a greater confidence in their ability to respond to suicide-related content after receiving the #chatsafe intervention, there was a lesser increase in young people’s intention to respond, and for a third of the sample, there was a decrease. In other words, possessing the confidence to communicate safely about suicide may not lead to actually engaging in a safe response or communication. This is not an uncommon finding in evaluations of mass media campaigns for suicide prevention, where raising awareness does not always translate to behavior change [49]. Alternatively, and perhaps more likely, the information provided by the #chatsafe intervention may have dissuaded young people from engaging in online conversations about suicide altogether, particularly if they were better able to assess the content that they come across as unsafe. The types of suicide-related content that participants reported seeing on social media suggest that they are mostly exposed to concerning content about this topic, and there is a chance that the information contained within the #chatsafe intervention empowered young people to disengage, block, or report that content rather than feeling the need to intervene.

The #chatsafe intervention provided general psychoeducation around suicide and digital literacy, and there was no heavy focus on encouraging young people to directly respond to suicide-related content online. A key message within the #chatsafe guidelines is for young people to check in with themselves and not feel the sole responsibility of engaging in conversations about suicide with someone that they are worried about. Despite often having the best intentions, some research suggests that young people who provide support about suicide or self-harm to others via social media report feeling worse themselves after that interaction [50]. However, the most preferred piece of content during this study included specific examples of how young people can approach a conversation about suicide, such as “it’s okay to feel that way” and “How do you feel about meeting for a coffee?” Although these are simple statements, guidance about what to say, or examples of words to use, likely address common fears about “saying the wrong thing” and may serve to protect those who would like to offer others support but feel ill-equipped to do so. This is a major gap in the current body of resources available to young people and one that future iterations of the #chatsafe intervention will attempt to address.

Most importantly, the #chatsafe intervention appeared to be safe and no adverse reactions were recorded. In addition, 97% of the participants reported that the content did not pose a risk to themselves, and 87% felt that it would not be a risk for others. That said, this study only retained approximately 60% of the
participants throughout the intervention period, and although retaining this proportion of young people in repeated measures studies is not uncommon [51,52], the findings should be interpreted with caution. While none of the participants indicated distress upon withdrawal, it is possible that some participants found the content overwhelming or unhelpful, which may be reflected in the finding that 40% of the participants did not believe that the content would be helpful in preventing future suicide deaths. This is unsurprising, as suicide is complex and unlikely to be prevented by a single intervention.

These findings support the potential for a social media intervention to play an important role in a broader postvention strategy, with a focus on disseminating age-appropriate and helpful information to young people. It has been recommended that after a suicide, postvention strategies aimed at mitigating suicide clusters need to be multifaceted and include a range of different approaches, including the monitoring of social media [28,30,32,33]. After a suicide has occurred, a social media intervention, such as #chatsafe, may result in safer online communication about suicide and subsequently act as a protective factor for young people in that online community [53]. Indeed, the outcomes of this study have already had practical implications for postvention responses delivered in real time and via social media. Since this study was conducted, the #chatsafe intervention material has been disseminated across communities in Australia (Western Australia, Victoria, and New South Wales) and New Zealand following a youth suicide. So far, at the time of writing, these interventions have reached ≥800,000 young people, and it is hoped that the #chatsafe content has contributed toward safer communication and the sharing of helpful information within those communities.

Strengths and Limitations
A key strength of this study was that it involved the delivery of a youth co-designed suicide prevention intervention shared within the environment in which young people are likely to encounter suicide-related information. Delivering interventions via social media makes them accessible, easily distributed, and relatively cost-effective [20]. It is also possible to reach large numbers of individuals in a short span of time. While young people at an elevated risk of suicide have historically been underrepresented in youth suicide prevention research [54], this study specifically recruited young people who had been impacted by a suicide or suicide attempt, a group known to be overrepresented in the suicide statistics. Furthermore, the study attracted a larger proportion of LGBTQIA+ (lesbian, gay, bisexual, transgender, queer, intersex, asexual) young people than the general population, another group who are disproportionately affected by suicide [55]. Despite the recruitment of young people within these groups, participants were predominantly cisgender females, and more work is required to understand the impact of the intervention on different groups of young people, particularly young males.

As this study was novel in its approach and the first of its kind, there are several learnings for future suicide prevention interventions delivered via social media. First, this was not a controlled study, and the changes observed cannot be directly attributed to the #chatsafe intervention. While this was a pilot study, a randomized controlled trial of the #chatsafe intervention commenced in November 2022 (Trial ID: ACTRN12622001397707). Second, this study did not collect information about the timing of suicide bereavement or exposure to a suicide attempt (other than it being within the past 2 years) nor did we collect information about the proximity to the suicide death or the subjective relationship with the deceased. This information is required to more thoroughly explore how the grieving process might impact the way in which #chatsafe content is perceived by young people. Third, although our questionnaire comprised measures and scales previously validated in other youth samples, they were not specifically designed to assess adherence to the #chatsafe guidelines and may not have adequately captured online behaviors and experiences relevant to the #chatsafe guidelines. This may account for the lack of predictor variables identified in our analyses. The ongoing randomized controlled trial using the #chatsafe intervention will use a new questionnaire that is tailored to measure adherence to the #chatsafe guidelines and more accurately address our research questions.

Previous work has identified that changes in willingness to intervene against suicide may be influenced by the type of exposure to suicide [56]. Participants in this study were eligible if they knew someone who had died by suicide and if they knew of a suicide attempt. Experiencing a suicide death versus knowing someone affected by suicide are qualitatively different experiences that are likely to impact the way one communicates about suicide and the way they are impacted by the communication of others [57]. Furthermore, the Circles of Vulnerability Model would argue that the degree of emotional impact felt by a suicide death is contingent upon 3 factors: geographical proximity, psychosocial proximity, and population at risk [11], yet little work has explored the role social media plays in determining proximity to suicide or in determining the closeness felt toward suicide-related content. Future research should seek to understand the differences in exposure and proximity (both online and offline) to develop and disseminate the most appropriate postvention material at the right time. Third, providing support to someone online is likely to be different from the offline context, and furthermore, recognizing and responding to risks may also be more challenging. It has previously been reported that perceptions of risk severity were a key factor influencing intent to intervene with a suicidal peer [58]. Observing others’ social media behavior is largely subjective, and this may make it a particularly challenging environment to offer support. Future research should explore the ways young people subjectively perceive distress or risk on social media so that interventions, such as the #chatsafe intervention, can best reflect the needs and wishes of young people.

Finally, this study found that young people are frequently exposed to harmful suicide-related content online. Although the guidance provided by the #chatsafe intervention aims to equip young people with the skills to communicate safely online about suicide, more information is needed to understand the impact of exposure (and at times, multiple exposures) on young people, particularly in relation to their own mental health. Further investigation of individual differences in the perception...
of risk and subsequent responses to suicide-related content will allow for more tailored intervention content for specific groups of young people in the future.

Conclusions
The findings of this study suggest that it is safe and acceptable to deliver a social media–based suicide prevention intervention to young people who have been exposed to a suicide or suicide attempt. The #chatsafe intervention social media content was received positively, and after exposure to the intervention, many participants reported a greater willingness to intervene against suicide, as well as increases in their perceived internet self-efficacy, confidence, and safety when communicating on social media about suicide. This was the first study to exclusively test the acceptability, impact, and safety of a suicide prevention social media intervention with a sample of recently bereaved young people. This study has provided preliminary evidence that #chatsafe is a safe and potentially efficacious intervention that could form part of future postvention responses and, as such, may have the potential to help reduce the risk of imitation or contagion after a suicide has occurred.

Acknowledgments
The authors would like to thank all young people who participated in this study. They would also like to thank their study partners, Portable. This study received funding from Suicide Prevention Australia. The #chatsafe project was funded by the Australian Government under the Department of Health’s National Suicide Prevention Leadership and Support Program. It also receives funding from the Future Generations Global and the William Buckland Foundation. LLS is funded by a Postdoctoral Fellowship from Suicide Prevention Australia. JP is funded by a National Health and Medical Research Council Investigator Grant (GNT1173126) and a Dame Kate Campbell Fellowship from the Faculty of Medicine, Dentistry, and Health Sciences at the University of Melbourne. JR is funded by a National Health and Medical Research Council Investigator Grant (ID2008460) and a Dame Kate Campbell Fellowship from the Faculty of Medicine, Dentistry, and Health Sciences at the University of Melbourne. SR was supported by a Career Development Fellowship from the National Health and Medical Research Council of Australia (GNT1158881) and a Dame Kate Campbell Fellowship from the Faculty of Medicine, Dentistry, and Health Sciences at the University of Melbourne. NTMH is funded by a Forrest Research Prospect Fellowship. The funders had no role in the study design, data collection and analysis, decision to publish, or manuscript preparation.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Report on change thresholds and predictors.
[DOCX File, 38 KB - humanfactors_v10i1e44535_app1.docx ]

References


Abbreviations

AE: adverse event
LGBTQIA+: lesbian, gay, bisexual, transgender, queer, intersex, asexual
OR: odds ratio
SAE: serious adverse event
TIDieR: Template for Intervention Description and Replication

©Louise La Sala, Jane Pirks, Charlie Cooper, Nicole T M Hill, Michelle Lamblin, Gowri Rajaram, Simon Rice, Zoe Teh, Pinar Thorn, Rifat Zahan, Jo Robinson. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 19.05.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Usability and Preliminary Efficacy of an Artificial Intelligence–Driven Platform Supporting Dietary Management in Diabetes: Mixed Methods Study

Kim Bul¹, BSc, MSc, PhD; Nikki Holliday¹, BSc, MSc; Mohammad Rashed Alam Bhuiyan²,³, BSc, MSc; Cain C T Clark¹,⁴, BSc, MSc, PhD; John Allen¹, BEng, MSc, PhD; Petra A Wark¹, BSc, MSc, PhD

¹Research Institute for Health and Wellbeing, Coventry University, Coventry, United Kingdom
²Centre for Trust Peace and Social Relations, Coventry University, Coventry, United Kingdom
³Department of Political Sciences, University of Dhaka, Dhaka, Bangladesh
⁴Warwickshire Institute for Diabetes, Endocrinology & Metabolism, University Hospitals Coventry & Warwickshire NHS Trust, Coventry, United Kingdom

Corresponding Author:
Kim Bul, BSc, MSc, PhD
Research Institute for Health and Wellbeing
Coventry University
Priory Street
Coventry, CV1 5FB
United Kingdom
Phone: 44 07392096816
Email: ac2658@coventry.ac.uk

Abstract

Background: Nutrition plays an important role in diabetes self-management. Web-based diabetes care, driven by artificial intelligence (AI), enables more personalized care.

Objective: This study aimed to examine the usability and preliminary efficacy of a web-based AI-driven nutrition platform to support people with diabetes and their carers in identifying healthy recipes, meal planning, and web-based shopping.

Methods: Diabetes UK signposted people with diabetes and their carers to the platform’s study-specific portal through its website, social media, and newsletters. A total of 73 adult participants with prediabetes or diabetes or their carers completed the baseline web-based survey. Of these 73 participants, 23 (32%) completed a web-based survey after 8 weeks of platform use. Web-based semistructured interviews were conducted with platform users (7/23, 30%) who agreed to be followed up and diabetes experts (n=3) who had nutrition and platform knowledge. The intervention consists of a web-based platform that incorporates AI to personalize recipes, meal planning, and shopping list experiences and was made available for 8 weeks. Baseline characteristics, satisfaction, system usability, and diabetes-related and general health indicators were assessed before and after using the platform for 8 weeks.

Results: Reductions in weight (mean difference 4.5 kg/m², 95% CI 1.0-12.0; P=.009; Cliff δ=0.33) and waist size (mean difference 3.9 cm, 95% CI 2.0-6.5; P=.008; Cliff δ=0.48) were found. Most of the participants (151/217, 69.6%) did not regularly use the platform and had low or very low engagement scores. However, the platform was perceived as accessible with no need for additional assistance (11/21, 52%), user-friendly (8/21, 38%), and easy to use (8/21, 38%), regardless of some usability issues. Saving recipes was the most popular feature, with 663 saved recipes.

Conclusions: This study indicated that the usability of the nutrition platform was well perceived by users and their carers. As participants managed their diabetes well, adding an education component would be specifically relevant for people less familiar with the role of diet in diabetes management. To assess the platform’s effectiveness in improving diabetes-related health indicators, controlled studies with a larger and more diverse participant sample are recommended.

JMIR Hum Factors 2023;10:e43959 doi:10.2196/43959
KEYWORDS
nutrition and dietetics; general diabetes; qualitative research; preventive medicine; web technology; self-management; diabetes; nutrition; deep learning; artificial; mobile phone

Introduction

Background
The prevalence of diabetes mellitus, one of the most burdensome noncommunicable diseases, has been rising as have its mortality rates and adverse societal and economic consequences [1-3]. Worldwide prevalence rates are as high as 9.3% (463 million people), with an expected increase to 10.9% (700 million people) by 2045 [4]. In 2018-2019, a total of 3.9 million people were diagnosed with diabetes in the United Kingdom. Diabetes can have severe health complications, including loss of eyesight, kidney disease, hypertension, heart failure, and diabetic feet, with rising economic costs for the National Health Service as high as £9.8 billion (US $12.4 billion) per year [5].

Type 2 diabetes is the most common form of diabetes, and it is associated with an unhealthy lifestyle in terms of physical exercise, nutrition, and weight. To manage diabetes in the long term, a combination of a healthy lifestyle and medication seems optimal. However, as individual factors (eg, comorbidities) can make glycemic control (glycated hemoglobin [HbA1c] levels) through medication challenging, a change in lifestyle is preferred as a first-step approach [6]. Under ideal circumstances, type 2 diabetes can be reversed, as demonstrated in a clinical trial where up to 46% and 36% of the patients had successfully reversed type 2 diabetes at 12 and 24 months after diagnosis, respectively [7]. In people with type 1 diabetes, there is a greater emphasis on counting carbohydrates and calories, but lifestyle management seems just as important to manage HbA1c levels and prevent health complications (eg, cardiometabolic risk) [8].

On the basis of the results of a systematic review and meta-analysis [9], lifestyle interventions seemed to reduce the incidence of type 2 diabetes and improve glycemic outcomes, anthropometric measures, physical activity, and energy intake across an ethnically diverse sample of adults at risk for developing type 2 diabetes compared with a control group. Focusing on lifestyle factors such as nutrition and physical activity does seem a promising avenue for prevention of type 2 diabetes in individuals considered to be at high risk.

Websites, mobile apps, artificial intelligence (AI) systems, serious games, automated calls and messages, and medical devices for diabetes prevention and care have gained popularity [10,11], especially during and after the COVID-19 pandemic, with the aim to improve care accessibility and self-management [12]. An overview of 15 systematic reviews showed that mobile health (mHealth) interventions can be effective in improving HbA1c levels, specifically for people with type 2 diabetes; however, the methodological quality of most of the reviews was limited [13]. A systematic review published later indicated that mHealth interventions have the potential to reduce weight, but the study findings, outcomes, and intervention durations were very heterogeneous [12]. A web-based education program offering support on nutritional management that was available for people with type 2 diabetes (or those with prediabetes) and their carers resulted in improved nutritional knowledge and people’s intentions to eat healthier and follow a healthy lifestyle [14].

Objectives
Most mHealth and health apps focus on monitoring blood glucose levels and have some educational components [12] (eg, by providing an AI-based embodied conversational agent that educates patients with type 2 diabetes about self-management) [15]. However, none of these digital apps used recipe exchange, meal planning, and web-based shopping features as catalysts for a healthy lifestyle and thereby diabetes management. This study aimed to examine the usability and preliminary efficacy of a web-based nutrition platform (using AI) that is freely accessible to anyone looking for support in identifying healthy recipes, meal planning, and web-based food shopping. The display of nutritional values is relevant for diabetes management, and this has been incorporated across all recipes on the AI-driven web-based platform. Although the application has not been specifically developed for people with diabetes and their carers, we hypothesized that the use of this nutrition platform will improve people’s general and diabetes-related health indicators, diet, and confidence regarding diabetes management.

Methods

Design
This mixed methods study had a pretest-posttest design. Quantitative data were derived from a web-based semistructured survey administered to people with diabetes (or their carers) and diabetes experts. Descriptive and inferential statistics (where applicable) were reported for participants’ general and diabetes-related health indicators before and after using the platform for 8 weeks. Platform data were captured with Mixpanel software [16] to assess real-time platform use.

Participants
The website of the charity Diabetes UK referred potential participants to the platform’s study-specific portal. People with type 1 or type 2 diabetes and those with prediabetes as well as those caring for someone with diabetes were eligible. Participants also had to be aged ≥18 years and have a good understanding of written and spoken English. A selection of the participants who consented to be approached for follow-up research were invited for semistructured interviews after 8 weeks of platform use. Diabetes experts were invited based on their nutrition and platform knowledge.

AI-Driven Web-Based Nutrition Platform
The platform uses AI to create an ecosystem for users that provides a better journey from recipe inspiration, meal planning, and food item shopping to (web-based) supermarket purchases. The AI software is informed by behavior (eg, viewed, shared,
shopped, and saved recipes), preferences (eg, health metrics, family size, diet, avoidances, favorite dishes, and like and dislike ingredients), and context (eg, weather, supermarket deals, user inventory, popular and trending recipes on the web, and food events). It generates personalized recipe suggestions, meal plans, shopping list items, and purchase options based on this information. Powered by deep learning and natural language processing using a natural language–based algorithm, the platform connects millions of data points about ingredients and their relationships to other ingredients, as well as recipe properties (eg, nutritional value, perishability, flavor, and category), including budget and availability across different supermarkets, to ensure good user availability within this ecosystem. Figures 1-3 present different features of the AI-driven web-based nutrition platform.

**Figure 1.** Screenshot of the platform showing the shopping list feature.
**Figure 2.** Screenshot of the platform showing the meal planner feature.

**Figure 3.** Screenshot of the platform showing the recipe feature.
Patient and Public Involvement Statement
This study did not involve patients or members of the public, but Diabetes UK represents people with diabetes and their carers, and thereby patient and public members as research participants.

Procedures
Participants were recruited from September 2020 to April 2021 through the charity Diabetes UK, which signposted participants to the platform’s study-specific portal. They could then sign up for access to the platform. Participants who provided informed consent for study enrollment were invited to take part in a web-based semistructured survey distributed through Qualtrics XM software [17] before and after using the platform for 8 weeks. A maximum of 2 electronic reminders (with a 1-week interval) were sent out where needed. Participants could indicate if they were happy be contacted for a web-based follow-up semistructured (up to 1 hour) or case study (up to 1.5 hours) interview, with the latter contributing to the creation of in-house personas. Diabetes experts from Diabetes UK were invited to participate in a semistructured interview. Data from both interview types have been merged and not presented separately owing to the similarity of identified key themes. Participants received a gift voucher of £10 (US $13.11; semistructured interview) or £20 (US $26.22; semistructured case study interview) to thank them for their time, whereas survey participants could participate in a prize draw (worth £55 [US $72.05]). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) checklist for cross-sectional studies has been used for this study reporting [18] (refer to Multimedia Appendix 1 [18] for the completed STROBE checklist).

Ethics Approval
The study was approved by the Faculty of Health and Life Sciences Ethics Committee at Coventry University (P109725) before study commencement (refer to Multimedia Appendix 2 for the original protocol for the study).

Assessments

Participant Characteristics and Device Use
Sociodemographic characteristics, disability, and smoking status were assessed through a web-based survey sent to participants after enrollment. The survey also asked about the kind of device and internet connection participants anticipated using to access the platform; for which purpose they currently use their device the most; and whether they were already using applications to monitor or improve physical activity, diet, or blood glucose levels. All questions provided the option “Prefer not to say.”

Computer Proficiency
The Computer Proficiency Questionnaire-12 (CPQ-12) was used at study commencement. Six domains—computer basics, printing, communication, internet, scheduling, and multimedia—were scored on a 5-point Likert scale ranging from 1=never tried to 5=very easily. The psychometric properties of the CPQ-12 have been shown to be comparable with those of the longer, 33-question version and were interpreted as excellent [19].

General Health Status
The EuroQol Visual Analog Scale (EQ-VAS) was administered before and after participants had used the platform for 8 weeks to measure their current general health status on a scale from 0 (the worst health imagined) to 100 (the best health imagined), with a score of 50 representing the population average [20]. The psychometric characteristics of the EQ-VAS have been described as satisfactory in people with diabetes [21].

Diabetes-Related Health Indicators
Height (measured in meters and centimeters or feet and inches), weight (kilograms or stones and pounds), waist circumference (centimeters or inches), blood glucose level (HbA1c; millimoles per mole), systolic and diastolic blood pressure (millimeters of mercury), and high-density lipoprotein (HDL) and total cholesterol levels (millimoles per liter) were self-reported and assessed before and after participants had access to the platform for 8 weeks. In addition, participants were asked when and by whom the last measurement took place. All questions provided the options “Prefer not to say” and “I don’t know.” Weight (kilograms) was divided by the square of the height (meters) to calculate BMI, which was categorized into underweight (<18.5 kg/m²), healthy weight (18.5-24.9 kg/m²), overweight (25.0-29.9 kg/m²), and obese (≥30.0 kg/m²).

Healthy Eating
Participants answered 8 questions about their diet before and after using the platform for 8 weeks, which were derived from the subscales Food frequency consumption and Food habits of an existing dietary questionnaire [22]. The questions provided an indication of eating habits relevant to people with diabetes and covered variation in diet, the type of snacks participants consumed, the consumption of sweets or cakes as well as fruits and vegetables, having or skipping breakfast, and water intake.

Confidence in Diabetes Management
Participants answered 3 questions focused on their confidence regarding diabetes management and meal planning before and after using the platform for 8 weeks. These questions (scored from 1=very unconfident to 10=very confident) were suggested by diabetes experts based on their experience of evaluating changes in people’s confidence in managing diabetes.

System Usability
The System Usability Scale assessed the platform’s usability [23]. This instrument has good psychometric properties and uses a 5-point Likert scale ranging from 1=strongly disagree to 5=strongly agree across 10 items. A total score of ≥68 is considered above-average usability [24].

Expectations and Satisfaction
Before they accessed the platform, participants answered 4 questions (scored from 1 to 10) on their expectations of using it, with a score of 1 representing very strongly disagree and a score of 10 representing very strongly agree. These questions were amended slightly to capture satisfaction with using the platform for 8 weeks. Participants also answered questions on how satisfied they were with individual platform elements as well as whether they learned anything new, encountered any
technical issues, and regarded the platform as user-friendly. Participants indicated how motivated they were to use the platform, whether they would recommend the platform to others, whether they would like to keep using the platform, and how the COVID-19 pandemic affected their use of the platform. Finally, they were asked to provide a general rating (ranging from 1 to 10) as well as recommendations on platform improvement.

**Platform Use**

User analytics were collected through Mixpanel software [16] by the platform developer. Data were collected on the number of platform sessions, saved recipes, recipes added to the meal plan, views of the shopping list, and engagement score for each participant who consented for their data to be shared with the research team. On the basis of platform use, the platform developer provided engagement scores classified into 5 groups (very heavy [score: ≥50], heavy [score: 20-49], medium [score: 10-19], light [score: 1-9], and none [score: 0]).

**Semistructured Interviews**

Semistructured interviews were conducted with platform users (7/23, 30%) and diabetes experts (n=3) to gain an understanding of platform usability and potential efficacy. Participants were interviewed individually over the web (via Microsoft Teams) or via telephone for up to 1.5 hours. Screen sharing was used to ensure understanding between the researcher and the participant when discussing specific aspects of the platform. Multimedia Appendix 3 presents the outline for the semistructured interview with participants, which was slightly amended for the diabetes experts. Audio recordings were transcribed verbatim. Thematic analysis was performed by 2 independent and experienced mixed methods researchers (KB and NH). By combining inductive and deductive coding using the interview outline as an initial coding frame, codes were created that were then clustered into themes and subthemes. Agreement was reached through comparison, discussion, and reflection [25]. Sample size was based on our expectation of reaching data saturation.

**Statistical Analysis**

Descriptive statistics were used to describe participants’ general and diabetes-related health indicators before and after using the platform for 8 weeks. To get an indication of the strength of evidence against the null hypothesis of no difference in indicators before and after using the platform for 8 weeks, the Wilcoxon matched pairs signed rank test was performed. In case of missing values, the whole pair was excluded from the analyses. This was supplemented by calculation of the Cliff δ effect size, which represents the probability of the superiority of 1 variable over the other, that is, the probability that a randomly selected observation from 1 group is larger than a randomly selected observation from another group, minus the reverse probability. The Cliff δ effect size ranges from −1 to 1, with 0 indicating stochastic equality of the 2 groups, where 1 indicates that 1 group shows complete stochastic dominance over the other group, and a value of −1 indicates the complete stochastic domination of the other group [26]. The values of δ = 0.15, 0.33, and 0.47 corresponded to small, medium, and large effect sizes, respectively [27]. Descriptive data analyses were conducted using SPSS software (version 26.0; IBM Corp) [28], whereas inferential data analyses were conducted using R Core Team [29] software (R Foundation for Statistical Computing).

**Results**

**Baseline Characteristics**

Participants’ sociodemographic characteristics at study start are presented in Table 1, and the recruitment flow is presented in Figure 4. Of the 73 participants who completed the baseline web-based survey, 23 (32%) filled in the survey after using the platform for 8 weeks. Of these 23 participants, 17 (74%) filled in the survey completely, 5 (22%) filled in >70%, and 1 (4%) completed 57% of the survey.

The survey participants were from England (61/73, 84%), Scotland (8/73, 11%), and Wales (4/73, 5%). More than half (42/73, 58%) of the participants had a diagnosis of type 2 diabetes. The average age of the participants was 59 (SD 11.1) years, with the youngest participant being aged 27 years and the oldest participant being aged 79 years; most of them were semiretired (28/73, 38%). Most of the participants were of White ethnicity (68/73, 93%), followed by Asian or Asian British (3/73, 4%). In terms of religion, participants stated that they were Christians (44/73, 60%), atheists (21/73, 29%), Hindus (2/73, 3%), Muslims (2/73, 3%), or that they practiced another religion (2/73, 3%). Participants were perceived as highly experienced in their computer use as indicated by a mean total CPQ-12 score of 27.5 (SD 3.5; range 14.5-30.0).

Most of the participants intended to access the platform on their iPhone or iPad (31/73, 43%), computer or laptop computer (21/73, 29%), or Android smartphone or tablet device (21/73, 29%). Most of the participants (33/73, 45%) had data access through Wi-Fi and the mobile phone network, whereas 30% (22/73) of the participants solely relied on Wi-Fi signals, and 25% (18/73) relied on a mobile internet connection. Participants mainly used their device to search for information (27/73, 37%), make calls and send SMS text messages or electronic messages (24/73, 33%), and use social media (10/73, 14%). Other uses included work, shopping, music, photography, and writing. The participants were already using a wide variety of apps (eg, Weight Watchers, Nutracheck, Noom, MyFitnessPal, and Myzone, FreeStyle Libre Sensor) on their device as well as sensors (eg, FreeStyle Libre Sensor) to monitor nutrition (37/73, 51%), physical activity (27/73, 37%), and blood glucose levels (20/73, 27%).
Table 1. Sociodemographic characteristics of survey participants at study start (n=73).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex, female, n (%)</strong></td>
<td>58 (79)</td>
</tr>
<tr>
<td><strong>Age (years), mean (SD)</strong></td>
<td>59.0 (11.1)</td>
</tr>
<tr>
<td><strong>Diabetes, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Type 1</td>
<td>10 (14)</td>
</tr>
<tr>
<td>Type 2</td>
<td>42 (58)</td>
</tr>
<tr>
<td>Prediabetes</td>
<td>12 (16)</td>
</tr>
<tr>
<td>Relative with diabetes</td>
<td>9 (12)</td>
</tr>
<tr>
<td><strong>Employment, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Full time or part time</td>
<td>29 (40)</td>
</tr>
<tr>
<td>Unemployed or unable to work</td>
<td>15 (21)</td>
</tr>
<tr>
<td>Other</td>
<td>28 (38)</td>
</tr>
<tr>
<td>Prefer not to say</td>
<td>1 (1)</td>
</tr>
<tr>
<td><strong>Education level, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>University: postgraduate or undergraduate</td>
<td>36 (49)</td>
</tr>
<tr>
<td>College</td>
<td>23 (32)</td>
</tr>
<tr>
<td>High school or secondary school</td>
<td>14 (19)</td>
</tr>
<tr>
<td><strong>Marital status: married (civil partnership) or cohabitating, n (%)</strong></td>
<td>54 (74)</td>
</tr>
<tr>
<td><strong>Number of people living in household, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>13 (18)</td>
</tr>
<tr>
<td>2</td>
<td>39 (53)</td>
</tr>
<tr>
<td>≥3</td>
<td>21 (29)</td>
</tr>
<tr>
<td><strong>Longstanding illness or disability, n (%)</strong></td>
<td>41 (56)</td>
</tr>
<tr>
<td><strong>Current smoking status, n (%)</strong></td>
<td>8 (11)</td>
</tr>
<tr>
<td><strong>Computer proficiency skills, mean (SD)</strong></td>
<td></td>
</tr>
<tr>
<td>Computer basics(^a)</td>
<td>4.8 (0.5)</td>
</tr>
<tr>
<td>Printing(^b)</td>
<td>4.3 (1.1)</td>
</tr>
<tr>
<td>Communication(^b)</td>
<td>4.9 (0.3)</td>
</tr>
<tr>
<td>Internet</td>
<td>4.7 (0.6)</td>
</tr>
<tr>
<td>Scheduling(^c)</td>
<td>4.4 (1.2)</td>
</tr>
<tr>
<td>Multimedia(^c)</td>
<td>4.3 (1.2)</td>
</tr>
</tbody>
</table>

\(^a\) n=70.  
\(^b\) n=71.  
\(^c\) n=72.
General and Diabetes-Related Health Indicators

There was no difference in participants’ reported general health status (mean difference [MD] −1.7, 95% CI −9.0 to 6.0; P=.61; Cliff δ=−0.05) before and after using the platform. However, weight (MD 4.5 kg/m², 95% CI 1.0-12.0; P=.009; Cliff δ=0.33) and waist size (MD 3.9 cm, 95% CI 2.0-6.5; P=.008; Cliff δ=0.48) were lower after 8 weeks of using the platform compared with baseline assessments. Most of the participants measured the diabetes-related indicators by themselves, but in some cases, either a medical professional or a nonprofessional else did so. Descriptive statistics of general and diabetes-related health indicators are reported in Table 2.

Table 2. Descriptive statistics of general and diabetes-related health indicators before and after using the platform for 8 weeks.

<table>
<thead>
<tr>
<th>Health indicator</th>
<th>Before platform use (n=73)</th>
<th>After 8 weeks of platform use (n=23)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Value, n (%)</td>
<td>Value, mean (SD)</td>
</tr>
<tr>
<td>General health (scale 0-100)</td>
<td>72 (99)</td>
<td>66.1 (20.0)</td>
</tr>
<tr>
<td>Height (m)</td>
<td>70 (96)</td>
<td>1.7 (0.1)</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>70 (96)</td>
<td>91.0 (22.6)</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>70 (96)</td>
<td>27.2 (6.7)</td>
</tr>
<tr>
<td>Waist size (cm)</td>
<td>57 (78)</td>
<td>101.0 (17.0)</td>
</tr>
<tr>
<td>Blood glucose level (HbA₁c, mmol/mol)</td>
<td>44 (60)</td>
<td>37.2 (31.1)</td>
</tr>
<tr>
<td>Blood pressure, systolic (mm Hg)</td>
<td>36 (49)</td>
<td>132.3 (13.7)</td>
</tr>
<tr>
<td>Blood pressure, diastolic (mm Hg)</td>
<td>35 (48)</td>
<td>78.2 (9.5)</td>
</tr>
<tr>
<td>HDL cholesterol level (mmol/L)</td>
<td>7 (10)</td>
<td>3.5 (1.7)</td>
</tr>
<tr>
<td>Total cholesterol level (mmol/L)</td>
<td>15 (21)</td>
<td>4.9 (0.7)</td>
</tr>
</tbody>
</table>

N/A: not applicable (height was not measured after 8 weeks of using the platform, given that this is a stable trait).

HbA₁c: glycated hemoglobin.

Median 42.0 (IQR 7.6-50.75).

HDL: high-density lipoprotein.

Median 3.6 (IQR 1.57-5.10).

HDL and total cholesterol levels are not reported after 8 weeks of using the platform, given that only 2 (9%) of the 23 participants reported these.
Healthy Eating

Dietary habits before and after using the platform for 8 weeks were comparable (Table 3).

Table 3. Descriptive statistics on aspects of healthy eating before and after using the platform for 8 weeks.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Before platform use (n=73), n (%)</th>
<th>After 8 weeks of platform use (n=22), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Participants’ diet is...</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Different every day</td>
<td>53 (73)</td>
<td>17 (77)</td>
</tr>
<tr>
<td>Different only sometimes during the week</td>
<td>11 (15)</td>
<td>4 (18)</td>
</tr>
<tr>
<td>Different during weekend days</td>
<td>2 (3)</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Very monotonous</td>
<td>7 (9)</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Snacking habits</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I snack</td>
<td>54 (74)</td>
<td>15 (68)</td>
</tr>
<tr>
<td>I do not snack</td>
<td>16 (22)</td>
<td>7 (32)</td>
</tr>
<tr>
<td>Prefer not to answer</td>
<td>3 (4)</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Having breakfast</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Always</td>
<td>51 (70)</td>
<td>18 (82)</td>
</tr>
<tr>
<td>Often</td>
<td>9 (12)</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Sometimes</td>
<td>8 (11)</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Never</td>
<td>5 (7)</td>
<td>2 (9)</td>
</tr>
<tr>
<td><strong>Consumption of sweets or cakes (number of times per week)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>12 (16)</td>
<td>2 (9)</td>
</tr>
<tr>
<td>Less than once a day</td>
<td>47 (64)</td>
<td>19 (86)</td>
</tr>
<tr>
<td>At least once a day</td>
<td>14 (19)</td>
<td>1 (5)</td>
</tr>
<tr>
<td><strong>Regularity of eating at least 2 portions (200 g) of fruit a day</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Always</td>
<td>26 (36)</td>
<td>9 (41)</td>
</tr>
<tr>
<td>Often</td>
<td>27 (37)</td>
<td>10 (46)</td>
</tr>
<tr>
<td>Sometimes</td>
<td>18 (25)</td>
<td>3 (14)</td>
</tr>
<tr>
<td>Never</td>
<td>2 (3)</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Regularity of eating at least 2 portions (200 g) of vegetables a day</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Always</td>
<td>33 (45)</td>
<td>12 (55)</td>
</tr>
<tr>
<td>Often</td>
<td>25 (34)</td>
<td>9 (41)</td>
</tr>
<tr>
<td>Sometimes</td>
<td>13 (18)</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Never</td>
<td>2 (3)</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Regularity of drinking at least 1 L of water a day</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Always</td>
<td>15 (21)</td>
<td>8 (36)</td>
</tr>
<tr>
<td>Often</td>
<td>25 (34)</td>
<td>3 (14)</td>
</tr>
<tr>
<td>Sometimes</td>
<td>22 (30)</td>
<td>8 (36)</td>
</tr>
<tr>
<td>Never</td>
<td>11 (15)</td>
<td>3 (14)</td>
</tr>
</tbody>
</table>

Confidence in Diabetes Self-Management

After using the platform for 8 weeks, participants felt most confident in meal planning (mean 6.0, SD 2.6; range 1-10) and making healthy food choices (mean 5.7, SD 2.6; range 1-10). They were least confident about their diabetes management before (mean 5.2, SD 2.6; range 1-10) and after using the platform (mean 5.4, SD 2.6; range 1-10).

System Usability

After using the platform for 8 weeks, participants reported a mean System Usability Scale index of 50.7 (SD 18.2; range 10-85), which indicated a below-average usability score. More than half of the participants (11/21, 52%) thought that they would not need assistance with using the platform. However, 43% (9/21) found the platform cumbersome to use, and 33% (7/21) found it unnecessarily complex. Of the 21 participants,
10 (48%) indicated that they would like to use the platform more frequently. Table 4 presents participants’ responses in more detail.

Table 4. System Usability Scale questionnaire scores for the platform (n=21).

<table>
<thead>
<tr>
<th>Statements</th>
<th>Disagreea, n (%)</th>
<th>Neutral, n (%)</th>
<th>Agreeb, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I think that I would like to use the platform frequently</td>
<td>7 (33)</td>
<td>4 (19)</td>
<td>10 (48)</td>
</tr>
<tr>
<td>I found the platform unnecessarily complex</td>
<td>8 (38)</td>
<td>6 (29)</td>
<td>7 (33)</td>
</tr>
<tr>
<td>I thought the platform was easy to use</td>
<td>7 (33)</td>
<td>6 (29)</td>
<td>8 (38)</td>
</tr>
<tr>
<td>I think that I would need assistance to use the platform</td>
<td>11 (52)</td>
<td>6 (29)</td>
<td>4 (19)</td>
</tr>
<tr>
<td>I found the various functions in the platform were well integrated</td>
<td>5 (24)</td>
<td>11 (52)</td>
<td>5 (24)</td>
</tr>
<tr>
<td>I thought there was too much inconsistency in the platform</td>
<td>6 (29)</td>
<td>13 (62)</td>
<td>2 (10)</td>
</tr>
<tr>
<td>I would imagine that most people would learn to use the platform very quickly</td>
<td>4 (19)</td>
<td>9 (43)</td>
<td>8 (38)</td>
</tr>
<tr>
<td>I found the platform very cumbersome to use</td>
<td>6 (29)</td>
<td>6 (29)</td>
<td>9 (43)</td>
</tr>
<tr>
<td>I felt very confident using the platform</td>
<td>7 (33)</td>
<td>8 (38)</td>
<td>6 (29)</td>
</tr>
<tr>
<td>I needed to learn a lot of things before I could get going with the platform</td>
<td>6 (29)</td>
<td>9 (43)</td>
<td>6 (29)</td>
</tr>
</tbody>
</table>

aScores 1 and 2 were combined and clustered under the heading of Disagree.
bScores 4 and 5 were combined and clustered under the heading of Agree.

Expectations and Satisfaction

Participants expected that the platform would support them primarily in making healthy food choices (mean 6.1, SD 1.7; range 1-10), planning meals more efficiently (mean 6.0, SD 1.9; range 1-10), diabetes management (mean 5.9, SD 1.7; range 2-10), and their food shopping experiences (mean 5.6, SD 2.1; range 1-10). After using the platform for 8 weeks, 18 (78%) of the 23 participants reported that the platform primarily supported them in planning meals more efficiently (mean 5.0, SD 2.7; range 1-10) and secondarily in diabetes management (mean 4.8, SD 2.0; range 1-9), making healthy food choices (mean 4.8, SD 2.2; range 1-9), and their food shopping experiences (mean 4.8, SD 2.6; range 1-10).

Most of these participants (17/18, 94%) indicated that they did not learn anything new while using the platform but found it to be easy to use (mean 5.6, SD 3.0; range 1-10). Of the 18 participants, 14 (78%) reported that they did not encounter any technical challenges. Half (9/18, 50%) of the participants indicated that they would recommend the platform to other people who have diabetes or are taking care of someone with diabetes. Participants scored 4.7 (SD 2.7; range 1-10) on the question regarding how motivated they were to use the platform, whereas 8 (44%) of the 18 participants said that they would not continue using the platform. The overall average rating of the platform was 5.2 (SD 3.2; range 1-10), and most of the participants (11/18, 61%) thought that the COVID-19–related restrictions did not affect the optimal use of the platform.

Platform User Statistics

Saving recipes was the most used feature across the platform, followed by adding recipes to the meal plan and viewing the shopping list. Survey participants who consented for their platform data to be shared and used the platform actively represented very heavy (3/33, 9%), heavy (5/33, 15%), medium (5/33, 15%), light (13/33, 39%), and none (7/33, 21%) categories of users (Tables 5 and 6).

Table 5. Platform user analytics per user group across total and survey sample (n=217).

<table>
<thead>
<tr>
<th>User pattern</th>
<th>Saved recipes, n</th>
<th>Frequency of viewing the shopping list, n</th>
<th>App sessions, n</th>
<th>Recipes added to the meal plan, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very heavy</td>
<td>663</td>
<td>50</td>
<td>313</td>
<td>91</td>
</tr>
<tr>
<td>Heavy</td>
<td>425</td>
<td>20</td>
<td>69</td>
<td>99</td>
</tr>
<tr>
<td>Medium</td>
<td>173</td>
<td>13</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>Light</td>
<td>132</td>
<td>33</td>
<td>15</td>
<td>21</td>
</tr>
<tr>
<td>None</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

aThe total amount of times 217 people saved a recipe.
Semistructured Interviews

Overview

Semistructured interviews were conducted during the COVID-19 pandemic. There was a general trend of people cooking at home more and trying out web-based recipes. The meal planner and web-based food shopping features became more popular during the COVID-19 pandemic because people wanted to restrict their outside shopping trips, and this held true regardless of low product stock and limited availability of delivery slots.

Theme 1: Usability

Participants described minor usability issues while using the platform. They indicated that the filter option for finding specific recipe content was not very visible:

[Diabetes expert 1] 
So I know that here you’ve got a filter system, but I think that’s not very obvious.

The layout of the diabetes-friendly community page on the platform was perceived as unclear. According to participants, there was an information overload on the home page and across platform features; in addition, measurement units, spelling, and ingredients were American and should be British:

I very rarely use a cup as a measurement because I don’t know what it is, I don’t know how big their cup is and this one has got yeast, so it’s important that you have got the right quantity of flour to yeast.

[Platform user 1]

Some of the participants found the recipe titles and cooking instructions on the platform (specifically the diabetes-friendly community page) unclear:

I would say most of the recipes that I’ve clicked on were quite good, but I saw this recipe for gnocchi, and I thought, oh yes, I quite like that, I’ll make that. And the instructions just didn’t make sense, total rubbish. I thought, there’s something gone wrong when they included that recipe in there...But yeah, I think you’ve got to check the recipes and make sure they’re at least readable, and understandable.

[Platform user 2]

Furthermore, information about portion size per serving was not displayed, it was not possible to change serving sizes for all imported third-party recipes, and there were synchronization issues among different devices and even among different users sharing and using the same web-based shopping list.

Participants perceived the platform as intuitive, user-friendly, and easy to use, mainly because it was primarily image based, and information was consistently and clearly presented across the platform:

I think because it’s so image based, it just makes it more engaging and easier to click, so I do think that helps. I know that’s not really...well, it’s part of the user experience, but I found it very easy to use generally, I’m just more about the...it’s easy to click and easy to add and all these things that I don’t think are terribly hard to use. [Diabetes expert 1]

Participants felt that there were sufficient instructions across the different platform features, which were easy to follow, although a platform user would have preferred more instructions:

Platform features were easy enough to use...yeah, even for somebody like me who is not a computer expert. [Platform user 2]

It took a few months before it was user friendly for me because it took me a while to understand it and I had to ask my daughter for some advice and she was able to do that. She said if you just press home and you go from there again, start again...and find the recipe I wanted, and press on it again. [Platform user 3]

Participants reported that navigation was easy and intuitive on their smartphones, but they preferred to use a larger screen (eg, tablet device or laptop computer) or printed recipes to follow instructions while cooking. A diabetes expert suggested that it might be easier to use the meal planner on a desktop computer because of the drag-and-drop functionality.

Theme 2: Perceived Usefulness

Participants perceived the platform as a useful starting point for people with diabetes. Others mentioned its usefulness for people with other health conditions, those who are less experienced cooks, and those who would like to eat healthier. The diabetes experts appreciated that the platform is community driven, and the indicators of popularity (eg, number of recipe likes and community members) were seen as useful by some. Participants reported that the platform offered a great deal of recipe inspiration and enabled people to conveniently collect and save recipes in 1 place.

Most of the diabetes experts were advocates of communicating glycemic index and load information, whereas the platform users mainly appreciated nutritional information such as carbohydrates, fat, and calories per serving. The diabetes experts thought that the meal planner was useful in offering structure as well as an overview of ingredients to support people in preparing a variety of meals throughout the week:

Table 6. Platform user analytics across a subsample of survey participants (n=33).

<table>
<thead>
<tr>
<th>Feature</th>
<th>Values, n(^a)</th>
<th>Value, median (IQR; range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Saved recipes</td>
<td>18</td>
<td>5.5 (2-17.25; 1-110)</td>
</tr>
<tr>
<td>Frequency of viewing the shopping list</td>
<td>13</td>
<td>1.0 (1-3; 1-13)</td>
</tr>
<tr>
<td>App sessions</td>
<td>6</td>
<td>6.5 (2.50-44.25; 1-144)</td>
</tr>
<tr>
<td>Recipes added to meal plan</td>
<td>9</td>
<td>3.0 (1-7.50; 1-14)</td>
</tr>
</tbody>
</table>

\(^a\)The total amount of times people who completed the survey (n=33) saved a recipe.
I think that’s all kind of useful stuff to help people meal plan better so that they’re not kind of trying to have new things every single day and that kind of recognition that you can repeat things throughout the week, that type of thing. [Diabetes expert 2]

They also thought that the web-based shopping feature was convenient and could save time even when the web-based shopping list was used to shop for items in a physical supermarket:

I just think the fact you can transfer to your shopping list makes it easier and on the go as well. People take their phone with them shopping, don’t they?...It’s great to see how it can easily be added to your shopping list in stores...That is a big asset, transferring it because it’s very...Recipes and writing a shopping list, it’s time consuming. [Diabetes expert 3]

Since like the New Year we’ve kind of did a big push to eat a bit better and also we found the shopping list and kind of meal planner bit really, really useful...I don’t think we’d kind of fully explored how we could integrate the planner and the shopping lists, that kind of element was quite a game changer on our part. [Platform user 4]

However, the platform users did not use the meal planner and shopping list features frequently, and some of them preferred paper-and-pencil methods because they perceived them to be a more flexible approach for people who are not very smartphone oriented:

At the moment I would say personally, no, but that’s because as I’ve explained that, you know, that it’s easier with pen paper, but I could see the facility would be useful for some people. I don’t knock the facility for those who work in that way and are very much attached to their smartphone, then I could see it being really useful tool. [Platform user 5]

Participants did not involve family members in their journey of using the platform but would definitely share it in case they thought it would be beneficial for them. However, some of the participants stressed the potential and enjoyment of sharing the platform with family members:

So me and my partner both have the app now, we share like a shopping list on there with each other and...Normally I will cook my own lunch through the app and then we’ll kind of cook a dinner together, normally from a recipe that we’ve both found on the app or that we’ve imported from wherever you know, from a website or from a cookery book. [Platform user 4]

Theme 3: User Experiences

Participants perceived the platform as attractive—nice, clean, fresh, and simple—but did not use it frequently. It provided a good variety of features, including the possibility to search, save, and like recipe content across different communities. The platform users expressed that the recipes were appetizing and seem to offer healthy food options for people with diabetes:

I liked the recipes because I hadn’t any idea what to cook for somebody that’s prediabetes, it’s something new to us, so I was very keen to see. [Platform user 3]

The drag-and-drop functionality used to transfer ingredients from the meal planner to the shopping list as well as the option to share the shopping list with relevant others were highly appreciated by some. In addition, some of the platform users indicated that the platform helped them to prevent food waste because they selected recipes on the basis of ingredients available at home.

Theme 4: Health Advice

Participants reported that the health score (Figure 5) displayed by the platform did not align with the traffic light labeling system and nutritional recommendations used in the United Kingdom.

According to the diabetes experts, it needs to be clear how the US-based health score was calculated and interpreted before presenting it to people with diabetes:

I felt that some of the recipes, the nutrient score, the health score didn’t necessarily reflect how healthy that recipe actually was. So I don’t know if that’s because of the nutrients that have been used. So like here, for example, this recipe gets a score of 5.9 and to me the only thing really that would be a negative is the salt. So yeah, I just wondered what the kind of justification was behind having all carbohydrates being a negative impact. I know they’ve got the fiber here as a positive so... [Diabetes expert 2]

One platform user with type 1 diabetes noticed that there was a difference between the nutritional values (specifically with regard to carbohydrates) displayed within the platform and those presented with the original recipes. A diabetes expert pointed out the potential risk of different nutritional analyses when American cups are widely used for measurement in British households. It is important to present information accurately and clearly because this can have direct implications on people’s blood glucose levels. The diabetes experts also thought that healthy recipe content review and approval by dietitians were necessary to ensure suitability for people with diabetes:

[People post anything on there that’s not vetted. I’ve got to be honest. That is a concern. It’s great it’s got choice and if you’ve got the knowledge, it is useful, but then, I’ve only spent 40 minutes with you and I’ve already clicked on quite a few things where we wouldn’t have it on our website. People with diabetes, both type 1 and type 2 are increased risk of cardiovascular disease and you’ve got a recipe with pure saturated fat. Of course I’m going to be concerned. [Diabetes expert 3]

Participants stressed the importance of not interpreting the platform as giving individual (medical) advice on what to eat, particularly when it has not been checked by a health care professional.
Theme 5: Potential to Improve Diabetes Management

The diabetes experts reported that labels such as diabetes friendly and low carb might be confusing for people with diabetes because a suitable diet is very individualized. They thought that it would be more important to focus on portion size, nutritional value, and type of carbohydrates. The majority of the participants felt that the platform offers a great variety of recipe inspiration and support for people with diabetes in meal planning and healthy eating habits:

People who are maybe looking for starting on that “how do I start to eat healthily,” it’s a good starting point, I would say definitely, and there’s a lot of recipes that means you could go with this for quite a number of weeks...I think most of the recipes that I’ve looked at are relatively straightforward. So, I don’t think they’re too overwhelming for people if suddenly they think how I am going to go with this. [Platform user 5]

So in terms of the planner and the shopping I think the planner itself will just kind of help people to...I mean it has the potential to help people kind of be thinking more proactively about what they’re eating rather than just making decisions on the spot or last-minute kind of decisions or whatever so it gives you that kind of forward thinking so that perhaps you can then decide about a range of different, you know...it potentially could help people to make healthier choices. [Diabetes expert 2]

Participants agreed that some education about what healthy eating entails and increasing knowledge of diabetes management were needed. The participants who were already quite knowledgeable about their diabetes and the role of nutrition felt that the platform did not contribute much to their diabetes management; however, they thought that it could be useful for those who have been recently diagnosed:

Again, I think the reason primarily for not is because I’ve lived with it for quite a long time, so a lot of that information is already in my head, but again going back to people who are newly diagnosed and I’d say in that, really thinking in terms of it from diagnosis and the first 12 months are the time when people feel that they’re under pressure, and so that system I feel probably again would be very helpful for people in that situation. [Platform user 6]

Some of the participants mentioned that the platform will only support diabetes management if people commit to changing their behavior and sustaining healthy habits, as well as have a good level of understanding about how food affects diabetes-related health indicators.

Theme 6: End-User Diversity

Participants reported that the platform had a universal approach suitable for anyone who wants to try out different recipes but...
would naturally attract people who speak English and are able to understand and use modern technologies. Because of the platform’s ease of use, this can include younger as well as older people, with the latter normally less experienced in using modern digital technologies. The diabetes experts mentioned that the platform has potential to involve the wider support system of people with diabetes (eg, friends, carers, and family members).

Participants mentioned that the platform displays some recipes from international cuisines (such as Asian and Indian but not African or Chinese), but it would not appeal to Asian, Black, and minority ethnic groups because it does not offer enough recipe variety:

It does strike me, and it may be that the people that are submitting the recipes are very city-centric, a bit London, affluent, digitally aware, disposable income...I don’t think it would, in the current format, appeal to a broad range of ethnicities or demographics. Some of the ingredients, for example. You’re looking at them. Elderflower syrup is on this one. [Diabetes expert 3]

[W]hether it’s getting people from all ethnicities to join, I don’t know. Plus obviously, it’s all in English. [Platform user 1]

Some of the participants indicated that the platform can never be fully inclusive but that this should be accepted because it offers enough variety and options for those who are interested. A diabetes expert indicated that feeling part of a community is more complex than sharing recipes over the web. Participants indicated that displaying a budget across recipes and including budget supermarket chains would potentially be helpful to reach communities with a lower socioeconomic status, specifically during the COVID-19 pandemic with more people being on furlough. Visual impairments and learning disabilities are quite common in people with diabetes; therefore, improving accessibility needs to be further considered.

Theme 7: Comparison With Other Apps and Platforms

Although there are other healthy recipe websites and applications available, participants reported that this platform is unique in offering a multicomponent approach by offering individual recipes, meal planning, and web-based shopping on 1 platform to support people with diabetes. The diabetes experts mentioned websites supporting education on diabetes and dieting and providing exercise applications that display nutritional analyses based on inserted data as well as applications to support counting carbohydrates and calories. Some dieting applications provide recipe videos in addition to written content, thereby making them accessible to a broader audience.

Theme 8: Recommendations for Improvement

Participants proposed several platform improvements regarding recipes, portion sizes, education and management, budget, peer support, tailored content, and layout (Multimedia Appendix 4).

Discussion

Principal Findings

This mixed methods study aimed to examine the usability and preliminary efficacy of an AI-driven web-based nutrition platform to support diabetes management. The survey and semistructured interview results showed that the platform was well received and primarily supported people with diabetes and their carers in identifying healthy recipes but less so in supporting meal planning and creating web-based food shopping lists. Although the diabetes-related health of most of the participants was largely stable, the platform was seen as attractive and a good starting point for recipe inspiration. The weight and waist circumference of participants tended to decrease after using the platform. However, because this is a small before-and-after study without a control group, it is not possible to conclude whether the improvements can be explained by actual platform use or other factors; therefore, these results should be interpreted with caution. High-quality robust trials are needed to examine its effectiveness on general and diabetes-related health outcomes.

Although recruitment was national, participation in the web-based surveys and semistructured interviews was low. This could be due to the COVID-19–related lockdown being associated with poorer mental health, especially among young women from ethnic minority groups who felt lonely and experienced prepandemic illness [30]. Therefore, they may not have prioritized research participation. Our sample seemed biased toward participants who were more willing to participate in study procedures [31], including primarily participants who are women, older, diagnosed with type 2 diabetes, nonsmoking, tech savvy, from a 2-person household, and of White British ethnic background. However, based on the semistructured interview results, participants expected that people who were less proficient with technology would not encounter any issues in using the platform because it is largely image based, easy to use and navigate, and only suffered from minor usability issues (eg, synchronization, American spelling, and measurements). Although the platform has been developed in line with accessibility and readability guidelines [32], the fact that participants preferred a larger screen to read instructions from while cooking and wrote down their shopping list using paper and pencil could mean that it was not experienced to be as accessible as it could be, considering potential vulnerabilities across the population with diabetes [33,34]. The diabetes experts recommended that visual impairments, ethnicity, and socioeconomic status should be considered when designing digital applications for patients with diabetes. Future applications should design readability and accessibility features that can be tailored to individual preferences, thereby increasing overall user experience [35].

In addition, participants reported a relatively high general health status and largely healthy eating habits, and their diabetes-related health indicators seemed to be within the normal range, apart from some of them being overweight (18/70, 26%) or having slightly elevated HDL cholesterol levels (22/70, 31%). Participants felt relatively confident about their ability to make
healthy food choices, plan meals efficiently, and manage their diabetes. This could partly explain why participants were only moderately motivated to use the platform and did not use the platform on a regular basis, as demonstrated by the platform user analytics.

On the basis of the semistructured interviews, participants indicated that the platform might be more relevant for people who have difficulties with managing their diabetes and are not aware of the impact of their eating habits (including nutritional values), had been recently diagnosed, or were on a waiting list to see a dietitian during the COVID-19 pandemic. The diabetes experts also felt that the platform would be beneficial for people with other long-term health conditions and people aiming for a planned approach to healthy nutrition and weight loss, which seems important in preventing diabetes. They were appreciative of the platform’s potential but also foresaw challenges in user-posted recipes as well as the accuracy of nutritional information and therefore felt that moderation by dietitians is needed. Indeed, this concern has also been raised across other web-based platforms providing specific nutritional information [36]. The diabetes experts also questioned whether the web-based platform would be able to realize actual behavior change over time. On the basis of a systematic review of web-based self-management programs for people with type 2 diabetes, it was clear that most of the studies (8/13, 62%) did not include a long-term follow-up [37]. Across the 5 studies that included follow-up assessments, only 1 (20%) assessed health outcomes beyond a 1-year follow-up. Future efficacy studies should include a longitudinal design to capture long-term effects in diabetes-related and general health outcomes and see whether improved lifestyle behaviors persist over time.

Despite our study not being representative of the population with diabetes and being too small to perform any kind of subgroup analyses (participants with high motivation vs those with low motivation), it still contributes to well-perceived platform experiences and its potential to support diabetes management. A variety of recruitment strategies, including using social media platforms to attract a younger population [38], should be explored in future studies to increase diversity in engagement and participation. This includes Asian, Black, and minority ethnic communities as a group with a high risk for diabetes [39,40], while also considering those with lower social economic status and comorbidities such as visual impairment [33,34].

Participants expected moderate levels of contribution of the platform toward making healthy food choices, planning meals efficiently, and managing diabetes. Indeed, the satisfaction scores indicated that the additional contribution of the platform regarding these aspects was below average, except for efficient meal planning. Although the meal planner was not frequently used, it could be that the use of recipes offered structure for cooking and meal preparation. High levels of self-efficacy are positively related to diabetes self-management behaviors [41], and this is likely to be the case in our study, given that the diabetes-related health indicators were mostly stable. Further research with larger samples will need to examine the role of motivation and self-efficacy in web-based interventions aimed at improving diabetes management [42].

Limitations
As participants were relatively healthy, and response rates were low, our sample and results may not adequately reflect the community of people with diabetes or only represent a small part of the community, namely participants who are healthier and may be more willing to cooperate or contribute to intervention research and its procedures. The reasons for high attrition on the survey could be explained by limited financial incentives, the absence of a prenotification to potential participants, and the use of a fully web-based survey instead of sending over the web as well as paper-based surveys [43].

Another reason could be that participants who did not use the platform on a regular basis or only used 1 aspect (eg, saving recipes) of the platform felt that they would not add value by filling in the survey about the platform, although we tried to preempt this by stating that regardless of how many times participants used the platform, completing the survey would still provide us with relevant insights into how to improve the platform. On the basis of the semistructured interviews, it seems that participants indeed showed curiosity in the beginning but did not maintain their interest over time, especially in the case of people who managed their diabetes relatively well and therefore felt that the platform was less relevant for them. The same pattern of engagement is seen in other web-based applications for chronic illnesses [44]. Although the reported usability issues were limited, high levels of engagement over a longer period of time with a web-based platform and study procedures remains challenging, and more research is needed in this area to determine which factors contribute to increased engagement, specifically for people with diabetes. Future research should anticipate higher attrition rates across participants who are chronically ill as well as older adult participants and account for this in their recruitment targets [45,46] while also considering different preferences in terms of delivery mode among participants to improve survey response rates.

In addition, Asian, Black, and minority ethnic communities, who normally display a higher prevalence rate of type 2 diabetes than European communities, were underrepresented in our study [39,40]. Therefore, as in the majority of other studies, our results cannot be generalized to the wider community of people with diabetes and need to be interpreted with caution owing to the study’s small and potentially biased sample. Robust and sample-diverse studies are needed to help inform subsequent priorities of research and applications in this area and draw conclusions more reliably across the whole community of people with diabetes.

Comparison With Prior Work
On the basis of platform user statistics, most of the participants (151/217, 69.6%) did not use the platform on a regular basis. This pattern is seen regularly across mHealth apps for diabetes where support tools are positively received, but the actual use is relatively low [47]. People who are managing diabetes using mHealth apps often need a reminder or push notification [47], which is not a default setting in this platform. This may need to be considered when the platform is offered to people with chronic health conditions to increase their engagement and
motivation to use features for recipes, meal planning, and creating web-based shopping lists. Apart from personalized recipe suggestions, using AI could improve engagement with such platforms and diabetes care [48]. Although participants were experienced in terms of computer use, and the majority (4273, 58%) owned a smartphone or laptop computer as demonstrated by the survey, it seemed from the semistructured interviews that participants did not embrace technology fully in daily life, as illustrated by some of the interview participants not taking their mobile phone with them while shopping or their preference to write down their shopping list using paper and pencil as well as to write down their favorite recipes in a notebook. This differentiates them from younger digital natives (defined as being born digital and therefore born in or after 1980 [49]) who share and synchronize their recipes and meal plans with relevant others, as demonstrated by an interview participant. However, this pattern contradicts the findings from a European cross-sectional survey study in which digital natives with type 1 diabetes used information and communication technology daily but not to support their diabetes care [50]. More research is needed on the barriers and facilitators to the use of technology across different digital age–specific groups.

Survey participants gave a relatively poor (or OK) usability rating to the platform, whereas the results from the interviews indicated that participants only experienced minor usability issues and felt that the platform was easy to use, user-friendly, intuitive, and contained clear instructions. The platform was mainly seen and used as a recipe inspiration platform in which recipe suggestions were prompted based on deep learning algorithms. The interview participants suggested that the platform would be useful as a starting point specifically for people recently diagnosed with diabetes (or other long-term chronic conditions) to eat healthier. This is particularly relevant, given the long waiting list of patients to be seen by dietitians or health care professionals and limited accessibility specifically during the COVID-19–related lockdowns.

Semistructured interviews with people with diabetes, their carers, and diabetes experts resulted in some recommendations to improve the platform. These mainly focused on recipes, administrator rights, portion size, education, budget, peer support and tailored content, and layout. An estimation of the cost per recipe will be useful for people affected by the current cost-of-living crisis and specifically support people from deprived areas where type 2 diabetes is more prevalent [34]. According to the diabetes experts and some of the platform users, it is important that nutritional information is validated and in line with the UK traffic light labeling system. Presenting nutritional information on calories and carbohydrates seems specifically relevant for people with type 1 diabetes, however, imprecise display of this information was seen as a potential barrier by an interview participant and the diabetes experts, especially given that other mHealth apps are available that display and calculate nutritional information more accurately and reliably [51]. Most of the participants acknowledged the importance of some education on the role of nutrition and physical activity levels in diabetes management, especially for people who had been recently diagnosed or lacked knowledge, before accessing the platform. This seems fruitful because knowledge, attitude, and practice seem to be positively related to glycemic control [52]. Although studies have suggested that education is an important aspect to improve existing mHealth apps in diabetes management, this has only been adopted by a few apps [15,47]. As the educational component was provided primarily in the nutritional values displayed in user-posted recipes, it is recommended to modify the web-based platform, which could be cocreated by people with diabetes and health care professionals, thereby using existing learning resources on diabetes management and minimizing costs [53].

On the basis of the diabetes experts’ input during the semistructured interviews, it should be noted that this platform does not in any way provide health or medical advice because nutrition in the context of diabetes is very individualized, and nutrition advice should always be sought from an appropriate health care professional. It was recommended to include some dietary monitoring of the recipes that are posted and shared across the community to ensure that all of them are in agreement with the UK national food-based dietary guidelines and traffic light labeling system [54]. This should be taken on board in further applications of the platform in the context of the management of long-term health conditions, such as diabetes.

Conclusions

The AI-driven, web-based nutrition tool was perceived as accessible and easy to use, with minimal usability issues. Several important recommendations for its improvement have been made, and the relevance of education on healthy eating for specific groups has been stressed. Most of our participants were quite knowledgeable and stable regarding the self-management of their diabetes. The potential of the platform’s meal planner and shopping list was acknowledged by the diabetes experts, but participants thought it was mainly useful for recipe inspiration. Diabetes experts indicated that the recipe content should be reviewed by experts to enable people with diabetes to maintain a reliable and healthy personalized diet.

Given that Asian, Black, and minority ethnic communities and other groups considered susceptible were underrepresented in this study, future research should deploy different recruitment strategies to involve a more representative sample of people with diabetes who could potentially benefit from this platform. This includes making a distinction among digital age–specific groups. Future applications should consider tailored accessibility and readability features to increase overall user experience. Although some reductions in weight and waist circumference were found, no causal inferences can be made because of the small sample size and the study’s pretest-posttest design. Longitudinal studies should examine the efficacy of web-based platforms regarding diabetes-related and general health outcomes. Cocreating solutions with people with diabetes and health care professionals and further development of AI technology have great potential to improve diabetes management in a more engaging and personalized manner. Motivation and self-efficacy are expected to play an important role, and theoretical underpinnings should be considered in intervention development and future studies.
Acknowledgments

The authors would like to thank the West Midlands Academic Health Science Network (WMAHSN) for supporting the performance of this study financially (PROC16.00019). The authors would also like to thank the platform developer for the technical support provided throughout the study and Diabetes UK for supporting participant recruitment and signposting its diabetes experts to be interviewed, as well as the participants for making the time and effort to contribute to this study.

Data Availability

The data set is available from the Figshare repository [55].

Authors' Contributions

KB wrote the initial draft of the manuscript and revised based on all author’s feedback. PAW wrote the funding proposal and thereby study design with support from KB and NH. KB and NH prepared, submitted and gained ethical approval. KB, NH and MRAB collected and analyzed qualitative and quantitative data. CCTC provided statistical advise and ran the inferential data analyses in R. All authors reviewed the final manuscript multiple times and provided intellectual contribution to this.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Reporting checklist for this cross-sectional study.
[DOCX File, 28 KB - humanfactors_v10i1e43959_app1.docx]

Multimedia Appendix 2

Original protocol for the study.
[DOCX File, 685 KB - humanfactors_v10i1e43959_app2.docx]

Multimedia Appendix 3

Outline for semistructured interview with participants.
[DOCX File, 17 KB - humanfactors_v10i1e43959_app3.docx]

Multimedia Appendix 4

Participants’ recommendations for platform improvements.
[DOCX File, 21 KB - humanfactors_v10i1e43959_app4.docx]

References


A Beta-Prototype Chatbot for Increasing Health Literacy of Patients With Decompensated Cirrhosis: Usability Study

Jessica Au1*, BMedSci, MD; Caitlin Falloon1*, BMedSci, MD; Ayngaran Ravi1, MBBS; Phil Ha2, MBBS; Suong Le2,3, MBBS

1School of Clinical Sciences, Monash University, Clayton, Australia
2Department of Gastroenterology and Hepatology, Monash Health, Clayton, Australia
3Monash Digital Therapeutics and Innovation Laboratory, Monash University, Clayton, Australia

*these authors contributed equally

Corresponding Author:
Jessica Au, BMedSci, MD
School of Clinical Sciences
Monash University
30 Wellington Road
Clayton, 3800
Australia
Phone: 61 431201056
Email: jessicaau7@gmail.com

Abstract

Background: Health literacy is low among patients with chronic liver disease (CLD) and associated with poor health outcomes and increased health care use. Lucy LiverBot, an artificial intelligence chatbot was created by a multidisciplinary team at Monash Health, Australia, to improve health literacy and self-efficacy in patients with decompensated CLD.

Objective: The aim of this study was to explore users’ experience with Lucy LiverBot using an unmoderated, in-person, qualitative test.

Methods: Lucy LiverBot is a simple, low cost, and scalable digital intervention, which was at the beta prototype development phase at the time of usability testing. The concept and prototype development was realized in 2 phases: concept development and usability testing. We conducted a mixed methods study to assess usability of Lucy LiverBot as a tool for health literacy education among ambulatory and hospitalized patients with decompensated CLD at Monash Health. Patients were provided with free reign to interact with Lucy LiverBot on an iPad device under moderator observation. A 3-part survey (preuser, user, and postuser) was developed using the Unified Acceptance Theory Framework to capture the user experience.

Results: There were 20 participants with a median age of 55.5 (IQR 46.0-60.5) years, 55% (n=11) of them were female, and 85% (n=17) of them were White. In total, 35% (n=7) of them reported having difficulty reading and understanding written medical information. Alcohol was the predominant etiology in 70% (n=14) of users. Participants actively engaged with Lucy LiverBot and identified it as a potential educational tool and device that could act as a social companion to improve well-being. In total, 25% (n=5) of them reported finding it difficult to learn about their health problems and 20% (n=4) of them found it difficult to find medical information they could trust. Qualitative interviews revealed the conversational nature of Lucy LiverBot was considered highly appealing with improvement in mental health and well-being reported as an unintended benefit of Lucy LiverBot. Patients who had been managing their liver cirrhosis for several years identified that they would be less likely to use Lucy LiverBot, but that it would have been more useful at the time of their diagnosis. Overall, Lucy LiverBot was perceived as a reliable and trustworthy source of information.

Conclusions: Lucy LiverBot was well received and may be used to improve health literacy and address barriers to health care provision in patients with decompensated CLD. The study revealed important feedback that has been used to further optimize Lucy LiverBot. Further acceptability and validation studies are being undertaken to investigate whether Lucy LiverBot can improve clinical outcomes and health related quality of life in patients with decompensated CLD.

(JMIR Hum Factors 2023;10:e42506) doi:10.2196/42506
KEYWORDS
chronic liver disease; chatbot; artificial intelligence; health literacy; acceptability

Introduction

Chronic liver disease (CLD) is a major global public health burden and results in 2 million deaths annually [1,2]. Decompensated CLD is a significant contributor to patient morbidity and mortality and is defined as an acute deterioration in hepatic function resulting in jaundice, hepatic encephalopathy, ascites, hepatoportal syndrome, or spontaneous bacterial peritonitis [3,4]. In 2012, the direct health care costs associated with the treatment of liver disease was estimated at US $448 million in Australia [5] and US $32.5 billion in the United States [6]. Lost productivity costs in Australia were estimated at US $4.3 billion in 2012, mainly from lost lifetime earnings due to reduced life expectancies and lower employment participation [3]. The World Health Organization [7] defines health literacy as “the achievement of a level of knowledge, personal skills and confidence to take action to improve personal and community health by changing personal lifestyles and living conditions. Thus, health literacy means more than being able to read pamphlets and make appointments. By improving people’s access to health information, and their capacity to use it effectively, health literacy is critical to empowerment.”

Poor health literacy has been demonstrated in patients with CLD, which may contribute to the high morbidity, mortality, and economic burden experienced by this specific chronic disease cohort [8,9].

Adherence to chronic disease treatment regimes has also been associated with health literacy, with adherence rates being 14% higher in patients with higher levels of health literacy [10]. In patients with liver cirrhosis, simple educational interventions increased patient’s disease knowledge by 26% [11].

Furthermore, the low levels of health literacy combined with high unemployment rates act as significant barriers for such patients to navigate complex health care systems and communicate with clinicians [12,13]. Studies have demonstrated an association between education level and CLD mortality, and this association was magnified for those with alcohol-related etiology [14,15]. The epidemiology of CLD is shifting away from chronic viral hepatitis toward lifestyle related etiologies including alcohol abuse and metabolic syndrome. This highlights the need for targeted interventions which address health literacy to improve self-management by reducing alcohol consumption, and addressing obesity, malnutrition, and sarcopenia [3]. Hepatic encephalopathy also impacts patient’s ability to understand health information, as it impairs executive function, problem-solving, and attention [9].

There are limited studies reporting the true prevalence of poor health literacy, its etiology, and the identification and management of potentially modifiable or preventable risk factors for poorer health literacy in decompensated CLD. Liver cirrhosis is a multisystem disorder, which is difficult for both clinicians and patients to optimize according to guideline-based management. There are high expectations placed on patients and carers to manage complicated medication regimes, lactulose self-titration, fluid and salt restriction, and nutrition optimization in the community. It has not yet been demonstrated whether improved health literacy is associated with increased patient self-sufficiency in these domains of cirrhosis self-management. Patients with CLD also have significant carer requirements, which negatively impacts the mental, physical, and social well-being of patients and caregivers [16]. This burden is further amplified in patients with hepatic encephalopathy and cognitive dysfunction [17].

A novel strategy to improve chronic disease patient engagement and self-management are artificial intelligence (AI) “Chatbots.” Chatbots are an emerging health care technology used for basic diagnostic or monitoring purposes in ambulatory settings [18,19]. An AI chatbot has the ability to use natural language processing (NLP) to decipher human language in order to retrieve relevant data using conversational algorithms [20]. This interactive user interface, which is intended to simulate a bidirectional conversation with a clinician aims to increase patient engagement and reduce information overload [21,22]. AI chatbots can be deployed through an omnichannel strategy: web-based, Facebook messenger, and mobile apps [23].

Recent studies have shown high levels of acceptance of health specific chatbots by users and physicians [21,24]. In psychiatry, a discipline where chatbots are more prevalent, they are used to screen for mental health disorders and are also capable of delivering cognitive behavioral therapy [25,26]. A key limitation of existing health care chatbots is their lack of human emotion [27,28] and limited focus on education [29] when they could be leveraged as a tool to improve health literacy among patients with complex chronic conditions. Providing targeted information to improve health literacy digitally could also help bridge the communication gap between patient and clinicians, while increasing patient autonomy [30].

A liver specific “Chatbot” that promotes CLD health literacy through an interactive conversational interface has not been reported in the literature. Our study aims to investigate whether a novel AI chatbot is an acceptable tool to provide health information to patients with decompensated CLD.

Methods

Study Design

We conducted a prospective mixed method study to determine patient usability of “Lucy LiverBot,” an AI chatbot designed and built by a clinical multidisciplinary team (MDT) at Monash Health, Australia using a no code platform provided by software developers Andi Chatterton and Mark Chatterton from inGeniousAI, an industry partner.

Lucy LiverBot

Overview

Lucy LiverBot is an AI chatbot developed by a MDT to deliver disease, medication, and nutrition-specific health information to patients with decompensated CLD (Figure 1). A key function
of Lucy LiverBot is the emphasis on health literacy and education [8]. Information is presented to patients through conversational scripts, visuals, and videos in English.

**Figure 1.** An example of a simple flow conversation with Lucy LiverBot which allows patients to type any questions they may have about chronic liver disease.

---

**Concept Development**

Lucy LiverBot was developed by a MDT from Monash and Austin Health in 2019 comprising a Hepatologist, a Liver Nurse Consultant, a Liver Pharmacist, and a Liver Transplant Dietitian. Each member of the MDT team was responsible for identifying a list of 10 questions commonly posed by patients with CLD in their area of subject matter expertise. Each MDT member was then responsible for creating the answers to these questions and for the veracity of the content. In-depth interviews and small focus groups were also conducted with 10 patients attending the Complex Liver Care Clinic, Monash Health—an ambulatory care program for adults with decompensated cirrhosis to validate the questions to be answered by Lucy LiverBot, identify any missing questions, and confirm a patient need for the product.

The Liver Pharmacist was trained as a superuser, built the Lucy LiverBot, tested the NLP, and was the primary data custodian of the frequently asked questions library. Questions that cannot be answered by Lucy LiverBot are notified by email to the Liver Pharmacist who consults the relevant subject matter expert and then builds the response to this new question in the backend. Although all patients used written cues, the device also allows for voice-to-text recognition. Emojis were also used to reduce the amount of written text.

**Usability Testing**

We conducted the study from when standardized tools to assess user’s satisfaction with the experience of using chatbots were unavailable. We did not use other usability tools such as the System Usability Scale or the Usability Metric for User Experience, as these tools were not developed to consider the conversational aspects which relate to a user’s interaction with a chatbot. Instead, we used a mixed methods approach to gather preliminary insights into a patient’s experience with Lucy LiverBot. Each participant engaged in a 1-time only testing session consisting of (1) a preuser testing survey to determine patient demographics, their baseline confidence levels managing their own health, their understanding of CLD nutrition and possible barriers to optimal health; (2) a user-testing survey to determine patient satisfaction and to analyze whether the user interface is patient friendly; (3) a postuser testing survey to determine overall satisfaction with the app and its likely use in the management of CLD. All 3 surveys were developed using the Unified Acceptance Theory Framework [31].
The presurvey was developed by JA and PH and reviewed by SL, a consultant gastroenterologist and hepatologist. The presurvey was a validated questionnaire based on patient reported measures of treatment burden—the “Patient Experience with Treatment and Self-Management (PETS)” [32]. The user testing phase was made up of 2 sections: participant use of the chatbot and a subsequent survey. Lucy LiverBot was preloaded onto iPads which were provided to participants who could ask any nutrition-related questions for approximately 15 minutes. Within this time frame, Lucy LiverBot would guide participants through specific conversation flows depending on the key words used in their initiating question. The user testing survey questions were developed with input from inGeniousAI as their experience in chatbot design and deployment provided a valuable insight into the strategies required for successful user testing. The post user testing survey occurred immediately after the user testing survey and was developed to assess overall patient acceptance and usability of the app. The surveys were self-administered but with the moderator present, documenting additional feedback verbally provided by the patients throughout the testing session.

Recruitment

Adult patients were recruited from liver clinics and inpatient wards at Monash Health, the second largest tertiary health care network in Australia. Monash Health provides 4.1 million episodes of care per year to a population of 1.5 million people. Given the intent to conduct a study, only 20 participants were recruited into this study. Study inclusion criteria were adults with decompensated CLD and capacity to provide informed consent. Clinical, laboratory, and imaging data were used to confirm patient CLD decompensation status. Decompensation was defined as per English as a second language criteria [33]. Participants were excluded from the study if they had greater than grade 1 hepatic encephalopathy at the time of consent, did not complete all components of the survey, or if they were unable to read, understand, or answer questions fluently in English.

Statistical Analysis

Baseline clinical and disease demographics including current state of liver cirrhosis, decompensation complications, and total burden of hospital admissions in the past 12 months were extracted from the patient medical record. Summary data are presented as means (SD), proportions, or median (IQR) depending on the data distribution. All verbal patient feedback was documented and captured by the moderator verbatim. Anonymized transcripts were uploaded onto NVivo (Lumivero) for Windows (version 1.3; Microsoft Corp) for data management and coding. Qualitative data were reviewed by 2 independent assessors and 2 sets of key themes were identified. A third independent assessor then synthesized these results to produce a final set of key themes. Illustrative quotes were reported to support themes.

Ethics Approval

Ethics approval (RES-19-461A) was granted by the Human Research Ethics Committee Monash Health and was carried out according to the National Statement on Ethical Conduct in Human Research (2018).

Results

Overview

The median age was 55.5 (IQR 46.0-60.5) years, with 55% (n=11) of them being female and 85% (n=17) of them being White (Table 1). The median BMI was 31.2 (IQR 22.6-36.55). Active alcohol consumption was cross referenced from both self-reporting and clinical documentation of alcohol being the confirmed etiology for CLD. No current alcohol intake was reported among 25% (n=5) of participants, 50% (n=10) of them were still current drinkers and 25% (n=5) of them had quit drinking alcohol.
Table 1. Baseline demographics (n=20).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years), median (IQR)</strong></td>
<td>55.5 (46.0-60.5)</td>
</tr>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>9 (45)</td>
</tr>
<tr>
<td><strong>BMI, median (IQR)</strong></td>
<td>31.2 (22.6-36.55)</td>
</tr>
<tr>
<td><strong>Ethnicity, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>17 (85)</td>
</tr>
<tr>
<td><strong>Smoking status, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Smoker or ex-smoker</td>
<td>11 (55)</td>
</tr>
<tr>
<td><strong>Alcohol intake, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Previous or current alcohol intake</td>
<td>15 (75)</td>
</tr>
<tr>
<td><strong>English fluency, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Fluent</td>
<td>18 (90)</td>
</tr>
<tr>
<td><strong>Highest level of education, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Less than high school</td>
<td>12 (60)</td>
</tr>
<tr>
<td>High school graduate</td>
<td>8 (30)</td>
</tr>
<tr>
<td>Tertiary</td>
<td>2 (10)</td>
</tr>
<tr>
<td><strong>Do you have difficulty reading and understanding medical information? n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>13 (65)</td>
</tr>
<tr>
<td>Sometimes</td>
<td>5 (25)</td>
</tr>
<tr>
<td>Usually</td>
<td>2 (10)</td>
</tr>
<tr>
<td><strong>Employment status, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Unemployed</td>
<td>19 (95)</td>
</tr>
<tr>
<td><strong>Carer, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>14 (70)</td>
</tr>
<tr>
<td><strong>Owns a device, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>18 (90)</td>
</tr>
<tr>
<td><strong>Cause of cirrhosis, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Alcohol</td>
<td>14 (70)</td>
</tr>
<tr>
<td>Viral</td>
<td>3 (15)</td>
</tr>
<tr>
<td>Other</td>
<td>3 (15)</td>
</tr>
<tr>
<td><strong>Time since diagnosis of cirrhosis, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Unsure</td>
<td>7 (35)</td>
</tr>
<tr>
<td>&lt;6 months</td>
<td>2 (10)</td>
</tr>
<tr>
<td>6 months to 2 years</td>
<td>4 (20)</td>
</tr>
<tr>
<td>2-4 years</td>
<td>4 (15)</td>
</tr>
<tr>
<td>&gt;4 years</td>
<td>3 (15)</td>
</tr>
</tbody>
</table>

In total, 35% (n=7) of participants reported having difficulty reading and understanding written medical information, despite 90% (n=18) of participants being fluent in English. A large proportion had not completed high school (n=12, 60%) and were unemployed (n=19, 95%); a majority (n=14, 70%) required a carer. The primary cause of liver cirrhosis in this patient group was alcohol (n=14, 70%), followed by viral (n=3, 15%) and other (n=3, 15%).

Of the 20 participants, 25% (n=5) of participants found it difficult to learn about their health problems and 20% (n=4) of them found it difficult to find medical information they could trust (Table 2). Although 65% (n=13) of them found it easy to
understand advice provided directly by their health care providers, and 25% (n=5) of them found it difficult to understand. In addition, 20% (n=4) of them found it difficult to find information on what foods they should eat to stay healthy and 45% (n=9) of them reported issues monitoring their eating and drinking habits. In addition, 70% (n=14) of them were bothered by feeling dependent on others for health care needs, with 35% (n=7) of them bothered when family or friends reminded them to do things for their health. Regarding emotional well-being, 50% (n=10) of them felt preoccupied by their self-care, with 55% (n=11) of them depressed about their CLD. A large proportion felt worn out by self-care (n=14, 70%) and were frustrated (n=15, 75%) with their health situation (Table 2).
Table 2. Difficulties experienced by patients in the self-management of decompensated cirrhosis.

<table>
<thead>
<tr>
<th>Survey questions</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Understanding medical information</strong></td>
<td></td>
</tr>
<tr>
<td>How easy or difficult has it been to learn about your health problems?</td>
<td></td>
</tr>
<tr>
<td>Easy</td>
<td>12 (60)</td>
</tr>
<tr>
<td>Neither easy or difficult</td>
<td>3 (15)</td>
</tr>
<tr>
<td>Difficult</td>
<td>5 (25)</td>
</tr>
<tr>
<td>How easy or difficult has it been to learn what foods you should eat to stay healthy?</td>
<td></td>
</tr>
<tr>
<td>Easy</td>
<td>14 (70)</td>
</tr>
<tr>
<td>Neither easy or difficult</td>
<td>2 (10)</td>
</tr>
<tr>
<td>Difficult</td>
<td>4 (20)</td>
</tr>
<tr>
<td>How easy or difficult has it been to find sources of medical information that you trust?</td>
<td></td>
</tr>
<tr>
<td>Easy</td>
<td>15 (75)</td>
</tr>
<tr>
<td>Neither easy or difficult</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Difficult</td>
<td>4 (20)</td>
</tr>
<tr>
<td>How easy or difficult has it been to understand advice from different health care providers?</td>
<td></td>
</tr>
<tr>
<td>Easy</td>
<td>13 (65)</td>
</tr>
<tr>
<td>Neither easy or difficult</td>
<td>2 (10)</td>
</tr>
<tr>
<td>Difficult</td>
<td>5 (25)</td>
</tr>
<tr>
<td><strong>Monitoring health behaviors</strong></td>
<td></td>
</tr>
<tr>
<td>How much of a problem has it been for you to monitor your health behaviors, for example, exercise, diet and medication adherence?</td>
<td></td>
</tr>
<tr>
<td>A little</td>
<td>12 (60)</td>
</tr>
<tr>
<td>Somewhat</td>
<td>3 (15)</td>
</tr>
<tr>
<td>Quite a bit</td>
<td>4 (20)</td>
</tr>
<tr>
<td>Not applicable</td>
<td>1 (5)</td>
</tr>
<tr>
<td>How bothered have you been by feeling dependent on others for your health care needs?</td>
<td></td>
</tr>
<tr>
<td>A little</td>
<td>5 (25)</td>
</tr>
<tr>
<td>Somewhat</td>
<td>7 (35)</td>
</tr>
<tr>
<td>Quite a bit</td>
<td>6 (30)</td>
</tr>
<tr>
<td>Not applicable</td>
<td>2 (10)</td>
</tr>
<tr>
<td>How bothered have you been by others reminding you to do things for your health, for example, take medications, eat healthy, schedule appointments?</td>
<td></td>
</tr>
<tr>
<td>A little</td>
<td>14 (70)</td>
</tr>
<tr>
<td>Somewhat</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Quite a bit</td>
<td>4 (20)</td>
</tr>
<tr>
<td>Not applicable</td>
<td>1 (5)</td>
</tr>
<tr>
<td><strong>Managing emotional well-being</strong></td>
<td></td>
</tr>
<tr>
<td>How often did your self-care make you feel preoccupied?</td>
<td></td>
</tr>
<tr>
<td>Rarely</td>
<td>8 (40)</td>
</tr>
<tr>
<td>Sometimes</td>
<td>8 (40)</td>
</tr>
<tr>
<td>Often</td>
<td>2 (10)</td>
</tr>
<tr>
<td>Not applicable</td>
<td>2 (10)</td>
</tr>
<tr>
<td>How often did your self-care make you feel depressed?</td>
<td></td>
</tr>
<tr>
<td>Rarely</td>
<td>8 (40)</td>
</tr>
</tbody>
</table>
Qualitative Results

Mental Health and Well-Being

Several participants identified improvement in mental health and well-being as an unintended benefit of Lucy LiverBot. Beyond providing disease specific information, the conversational nature of the chatbot appealed to many as it provided a well-received reminder to maintain habits conducive to their well-being and health habits.

Maintaining wellbeing.

Keep checking up on you.

It was 1 participant who suggested additions that could be incorporated into the chatbot to specifically focus on the mental health of users.

Something for anxiety and depression could help, particularly being able to write a journal.

Lucy LiverBot was overtly identified as a potential “companion” by patients with CLD by providing a sense of social connection for patients who are socially isolated.

You can talk like you are talking to somebody else like a friend.

Timing of Chatbot Implementation

A common theme that emerged from participants was that the use of Lucy LiverBot may depend on the timing of its implementation in the patient’s disease progress. Patients who had been managing their liver cirrhosis for longer periods of time identified that they would be less likely to use Lucy LiverBot at later stages of CLD, but that it would have been useful at the time of their diagnosis.

Telling stuff I already know so not that useful.
Brilliant, can help many people, for young people.

Reliable Source of Medical Information

Lucy LiverBot was generally perceived as a reliable and trustworthy source of information as it was produced by medical professionals in the field of CLD. Participants recognized their potential to provide a trusted reference for nutritional information, rather than resorting to the internet.

Having information that is not conflicting.
Very informative. Would explain everything if I didn’t know anything about cirrhosis.
It’s quick and very simple to use. I like how I can ask questions as soon as they arise rather than wait for an appointment or google world wide.

Discussion

Principal Findings

There is a clinical urgency for cost-effective and scalable interventions that address the poor health literacy of patients with CLD [34] in order to improve patient engagement and self-management of this complex condition [35,36]. Lucy LiverBot was well received by participants and the results suggest that it could provide targeted CLD information via an engaging channel. Participants were actively engaged while using Lucy LiverBot throughout the in-depth user testing.
process, which took approximately 1 hour. We were also able to capture users from a variety of age groups and at different stages of their disease process which allowed us to determine at what stage of CLD Lucy LiverBot would be most useful. In addition, our extensive testing process ensured that all available chat flows were tested and NLP continued to improve with each consecutive patient.

Our results highlighted key barriers faced by patients with CLD which have the potential to impact their health outcomes—understanding health information, monitoring health behaviors, managing emotional well-being, and multidisciplinary communication. Lucy LiverBot has been specifically designed to assist patients with the understanding of health information and the monitoring of their health behaviors. It is also hoped that a centralized digital device designed by the MDT will help bridge the communication gap between patients and clinicians.

Many participants identified that the conversational tone and companion-like nature of the chatbot was one of its key strengths. Lucy LiverBot’s ability to engage with users provided a social platform for them to ask concerns and may have the potential to extend its disease specific content to directly address mental well-being and provide a sense of social connection. By addressing these identified barriers, Lucy LiverBot has the potential to fill a gap in the provision of health care to this group of complex chronic disease patients. Further validation studies are required to determine whether Lucy LiverBot as an intervention would prevent clinical outcomes such as readmission related to decompensated CLD.

It would be important to continue monitoring the performance of Lucy LiverBot after its launch to identify any errors in NLP so that necessary adjustments can be made. The NLP feature in Lucy LiverBot is basic and further advancements in this technology will be required to improve future iterations capable of providing an even more engaging user experience. The user testing allowed us to gauge how patients were most likely to phrase questions which allowed us to alter recognized terms. This was evident as Lucy LiverBot was unable to recognize some patient questions during the user testing stage if they were not worded in a similar manner to the initial input options. Unfortunately, the chatbot is currently only available in English, which limits its generalizability and scalability for participants from Culturally and Linguistically Diverse communities. This precluded some patients from participating in this study, however this is a technical limitation of NLP in general, rather than of Lucy LiverBot specifically. Ideally, future versions of Lucy LiverBot will be available in multiple languages.

The study was limited by a paucity of research on health chatbots, which made it difficult to determine the sample size required to adequately power the study and the ideal study design to assess patient usability. Our small study population allowed preliminary information to be obtained regarding the usability of Lucy LiverBot and its potential to act as an educational tool for patients with CLD. However, future studies with larger cohorts of patients will be required to definitively demonstrate Lucy LiverBot’s ability to improve health literacy and health outcomes. It is likely that solutions such as Lucy LiverBot will require frequent cycles of iteration and user testing to become maximally effective. This will lend more insight into what features are most beneficial within Lucy LiverBot and whether patients will be committed to using it for an extended period of their own volition beyond a study context. There was a potential selection bias as participants who agreed to join the study may be more motivated to improve their health and more likely to engage with Lucy LiverBot. In addition, response bias may have played a role in the study as participants completed the survey while investigators were in the room for technical support. This may have influenced patients to select answers that they believed were more acceptable. To remove any potential responder bias due to a perceived impact on their care, participants were assured that the results from the study would not be viewed by their treating team. We also trained final year medical students to conduct the testing rather than physicians to reduce the perceived power imbalance between participants and interviewers. The study also did not assess the stage of hepatic encephalopathy in participants. In future studies, we plan to assess this both at baseline and longitudinally to further delineate the effect of hepatic encephalopathy on a patient’s ability to remember health specific information.

The efficacy of novel digital health interventions such as Lucy LiverBot, which lack a formal evaluation framework akin to pharmacotherapy and device trials, would benefit from a multidisciplinary evaluation strategy tailored to the specific study end point. We conducted the study from when standardized tools to assess user’s satisfaction with the experience of using chatbots were unavailable. If we were to repeat the study again we would leverage new tools such as the Chatbot Usability Scale. An assessment of human computer interactions will also be required to determine the real-world patient usage patterns of Lucy LiverBot. A randomized controlled trial would be the ideal format in determining whether Lucy LiverBot is effective in improving health literacy and reducing hospital readmission. However, as ambulatory medical care models become increasingly multidisciplinary, it may become difficult to delineate which arms of the multimodal health care model are responsible for changes in clinical outcome. For example, if an improvement in admissions for hepatic encephalopathy were to be observed, this could be attributed to the increased communication with clinical staff through chatbot alerts, the health education provided by the chatbot, or perhaps an improvement in other indices such as nutrition and adherence. Further prospective studies based on the principles of implementation science are warranted to assess the benefits that Lucy LiverBot may provide to clinical end points such as readmission rate, morbidity, quality of life, and clinic attendance. A longitudinal component of the study should be established whereby participants are tracked throughout their disease progression and compared to those who did not use a health chatbot to determine whether Lucy LiverBot prevented hospital readmissions and led to improved patient outcomes. Such studies will need to perform costings analyses and assess long-term patient participation, and adherence to digital health care models.

Conclusions

Our study identified barriers to health care provision and found that Lucy LiverBot was well received by patients with
decompensated chronic liver disease. Lucy LiverBot can specifically address these barriers and be introduced as a potential educational intervention to address the impact of poor health literacy on disease outcomes and a health related quality of life. Further validation studies are required to demonstrate the potential for Lucy LiverBot to improve patient engagement and self-management and its use as an engagement tool with multidisciplinary teams.

Acknowledgments

We are incredibly grateful to our patients for taking the time to contribute to research into the use of chatbots in chronic liver disease.

Authors’ Contributions

JA, CF, AR, PH, and SL were all involved in data curation, conceptualization, formal analysis, investigation, methodology, project administration, writing-original draft, and writing-reviewing and editing. SL was also involved in funding acquisition, validation, and supervision. All authors contributed to the interpretation of the data, drafting of the manuscript revisions, and had final responsibility for the decision to submit for publication.

Conflicts of Interest

SL received funding from an investigator-initiated research grant provided by Norgine Pharmaceuticals.

References

5. Gastroenterological Society of Australia/Australian Liver Association. The economic cost and health burden of liver diseases in Australia. Deloitte Access Economics. 2013. URL: https://static1.squarespace.com/static/50f08044b007d5a9abe0a5/t/53321a3ee4b0f967eb0c7e5/1395792586844/gesa2013_revised%5B1%5D.pdf [accessed 2022-03-01]


Original Paper

eHealth Literacy and Patient Portal Use and Attitudes: Cross-sectional Observational Study

Nikita Deshpande¹, MD; Vineet M Arora², MAPP, MD; Hanna Vollbrecht³, MD; David O Meltzer⁴, MD, PhD; Valerie Press², MD, MPH

¹Pritzker School of Medicine, University of Chicago, Chicago, IL, United States
²Section of General Internal Medicine, Department of Medicine, University of Chicago, Chicago, IL, United States
³Department of Medicine, Brigham and Women’s Hospital, Boston, MA, United States
⁴Section of Hospital Medicine, Department of Medicine, University of Chicago, Chicago, IL, United States

Corresponding Author:
Valerie Press, MD, MPH
Section of General Internal Medicine
Department of Medicine
University of Chicago
MC 2007, Rm B224
5841 S Maryland Ave
Chicago, IL, 60637
United States
Phone: 1 773 702 5170
Email: vpress@bsd.uchicago.edu

Abstract

Background: Throughout the COVID-19 pandemic, patient portals have become more widely used tools of patient care delivery. However, not all individuals have equivalent access or ability to use patient portals.

Objective: The aim of this study is to evaluate the relationships between eHealth literacy (eHL) and patient portal awareness, use, and attitudes among hospitalized patients.

Methods: Inpatients completed patient portal surveys; eHL was assessed (eHealth Literacy Scale). Multivariable logistic regression analyses adjusted for age, self-reported race, gender, and educational attainment were completed with significance at P<.006 (Bonferroni correction).

Results: Among 274 participants, most identified as Black (n=166, 61%) and female (n=140, 51%), mean age was 56.5 (SD 16.7) years, and 178 (65%) reported some college or higher educational attainment. One-quarter (n=79, 28%) had low eHL (mean 27, SD 9.5), which was associated with lower odds of portal access awareness (odds ratio 0.11, 95% CI 0.05-0.23; P<.001), having ever used portals (odds ratio 0.19, 95% CI 0.10-0.36; P<.001), less perceived usefulness of portals (odds ratio 0.20, 95% CI 0.10-0.38; P=.001), and lower likelihood of planning to use portals in the coming years (odds ratio 0.12, 95% CI 0.06-0.25; P<.001). As time through the COVID-19 pandemic passed, there was a trend toward increased perceived usefulness of patient portals (53% vs 62%, P=.08), but average eHL did not increase through time (P=.81).

Conclusions: Low eHL was associated with less awareness, use, and perceived usefulness of portals. Perceived usefulness of portals likely increased through the COVID-19 pandemic, but patients’ eHL did not. Interventions tailored for patients with low eHL could ensure greater equity in health care delivery through the COVID-19 pandemic.

(JMIR Hum Factors 2023;10:e40105) doi:10.2196/40105

KEYWORDS

health literacy; patient portal; COVID-19; health technology; inpatients; digital health literacy; awareness; use; engagement; attitudes; hospitalized patients; access; accessibility; perception; health care delivery
**Introduction**

### Increasing Relevance of Patient Portals

Patient portals are increasingly important tools for providing patient care [1-6]. They are used to schedule appointments, view results, request medication refills, and communicate with health care professionals [1,6]. Recently, patient portals have become increasingly salient, playing a vital role in vaccine distribution [5], COVID test result notification [3], and maintenance of care [2,4] virtually through disruptions in service through the COVID-19 pandemic.

### Disparities With Portal Use and Access

As with all new technology, it is vital to assess how existing health and health care disparities are impacted by the growing use of these patient portals. Prior studies have found that some populations, such as individuals who identify as Hispanic or Black and individuals with lower educational attainment are less likely to access patient portals [7,8]. Furthermore, older patients have been found to be less likely to enroll in patient portal programs [1,7]. The digital divide describes disparities in individuals’ access to and capabilities to use technology and differences in outcomes when using technology. Key determinants of the divide have been shown to include age, educational attainment, and socioeconomic status [9-11].

### eHealth Literacy

*eHealth literacy (eHL)* characterizes patients’ ability to find, comprehend, and evaluate health information from electronic sources [12]. Patients with lower eHL have been found to use the internet less often and to be less likely to search for health information [13]. The *eHealth Literacy Scale* (eHEALS) has been validated in diverse patient populations and is a frequently used measure of eHL [14,15]. Similar to other tools, it has limitations, including lacking items measuring skills and comfort with navigating social media sites and peer support forums [16,17].

### Study Aim

Past study of patient portals has focused on the outpatient setting, but understanding portal and use and attitudes among admitted patients is also important and may capture a more impaired, high-risk patient population. To our knowledge, the relationship between eHL and patients’ engagement with portals has not been characterized among general medicine inpatient populations. This study aimed to characterize how age, self-reported race, and eHL were associated with portal awareness, use, and perceptions among adult inpatients at UChicago Medicine.

### Methods

#### Study Design and Participant Population

Inclusion criteria were being 18 years or older, speaking English, and being admitted to a general medicine service. Patients who lacked decisional capacity due to altered mental status or some other conditions were excluded. The recruitment occurred during the daytime for eligible patients at any time during their hospitalization. Patients provided their consent to the trained research assistants who recruited patients and filled out demographic, eHL, and survey data on access to and use of technology, including patient portals.

### Ethical Considerations

This cross-sectional, observational survey was completed as a part of a larger quality of care study approved by the University of Chicago Biological Sciences Division institutional review board (#IRB16-0763).

#### Data Collection and Analysis

According to previous literature, low eHL was considered <24 [13]. To evaluate technology use and access, participants were asked if they owned technological devices, if they had wireless internet at home, and how frequently they accessed the internet. To assess patient portal awareness and use, the participants were asked if they were aware of access to a patient portal and had used a patient portal in the past. To evaluate patient portal attitudes, participants were asked how confident they were in their ability to use a portal, how useful they believed a portal was, and how likely they would use a portal in the next year (Multimedia Appendix 1). The validated 8-item eHEALS tool assessed eHL [14]. The eHEALS tool asks patients about their ability and confidence in finding and discerning health information on the internet (Multimedia Appendix 2) [18]. Surveys were administered either in-person or over the phone. Cases with missing data were omitted.

Descriptive statistics included means, SDs, and proportions. Bivariate chi-squared analyses were conducted. Multivariable logistic regression analyses were performed to determine the differences in patient portal use and attitudes, adjusted for eHL (binary), age (binary, <65 vs ≥65), gender (binary), self-reported race (White, Black, and others), and education (high school diploma or less vs some college or more). A *P*<.006 defined statistical significance based on Bonferroni correction [19]. STATA (version 15.1; StataCorp LLC) was used for all analyses.

### Results

#### Study Design and Participant Population

#### Study Enrollment

From January 11, 2020, to August 3, 2021, a total of 2795 patients were screened and 1957 (70%) were eligible. Of those eligible, 274 participants (14%) were enrolled and completed the survey. Demographic data of those who refused, were discharged before the approach, or were not available during the approach were not recorded. Overall, 93% (255/274) of surveys were administered over the phone.

#### Participant Characteristics

The mean age was 56.5 (SD 16.7) years. The majority of participants identified as Black (166/274, 61%) and female (140/274, 51%). Sixty-five percent (178/274) reported some college or higher educational attainment, 33% (90/274) reported at most a high school education, and 2% (6/274) did not know or declined to say (Table 1). The majority of participants did...
not know or declined to provide annual household income (190/274, 69%).

Table 1. Distributions and odds ratios for bivariable and multivariable logistic regressions predicting portal awareness, use, and attitudes.

<table>
<thead>
<tr>
<th>Data Collection and Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Participant Technology Ownership, Use, and eHL</strong></td>
</tr>
<tr>
<td>Most participants owned at least 1 technological device (260/274, 95%), had Wi-Fi access at home (219/274, 80%), and used the internet several times per day (192/274, 70%). Overall, 28% (79/274) of participants had low eHL (range 8-40; mean eHEALS score 27, SD 9.5).</td>
</tr>
</tbody>
</table>

| Associated Factors of Portal Use and Attitudes |
| Low eHL (odds ratio [OR] 0.11, 95% CI 0.05-0.23; \( P < .001 \)) and identifying as Black (OR 0.18, 95% CI 0.06-0.55; \( P = .002 \)) were associated with lower odds of being aware of access to a portal (Table 1). Low eHL (OR 0.19, 95% CI 0.10-0.36; \( P < .001 \)) was associated with lower odds of ever using a portal. Low eHL was associated with less perceived usefulness of patient portals (OR 0.20, 95% CI 0.10-0.33; \( P = .001 \)) and low eHL (OR 0.12, 95% CI 0.06-0.25; \( P < .001 \)) was associated with not planning to use portals in the coming year. The most common reasons why participants had not used portals in the past year included being unaware of their access (68/274, 25%), unable to set it up (27/274, 10%), and feeling it would not improve their health care experience (17/274, 6%). |

| Changes in Portal Attitudes and eHL Through Time |
| Data were separated into quartiles based on survey administration date to evaluate trends over time. As time passed through the COVID-19 pandemic, there was a trend toward increased perceived usefulness of patient portals (53% [Q1] vs 62% [Q4]; \( P = .08 \)), but average eHL did not increase through time (\( P = .81 \)). |

**Discussion**

**Principal Findings**

Low eHL was associated with less portal awareness and past use. It was additionally associated with more negative patient portal attitudes, including less perceived usefulness and less likelihood of planning to use a portal in the next year. Older age was also associated with lower odds of planning to use a portal in the future. While the COVID-19 pandemic resulted in trends toward increased perceived usefulness of portals through time, patients’ eHL did not increase through the pandemic, suggesting that the patients were not empowered to better use digital tools as the pandemic progressed.

These findings extend previous studies that the digital divide is shifting from a disparity in access to a disparity in digital capabilities (as measured by eHEALS) [20-22]. More than 90% (n = 260) of patients in our sample had access to at least 1 technological device, but only two-thirds (n = 195) had adequate eHL. Furthermore, this study extends the findings of correlation between eHL and patient portal use previously reported among outpatients and organ transplant recipients to a hospitalized, urban, predominantly Black general medicine population [23]. Studying eHL and portal attitudes among inpatients captures individuals during the unique stressor of hospitalization and patients who may not engage with outpatient medicine and may otherwise be missed. Future efforts to increase patient utilization of portals likely needs to shift from simply increasing access to the internet to other interventions such as increasing awareness of the usefulness of portals and interventions to assist patients with portal use, particularly among patients who are older and have low eHL [21].
Addressing Low Portal Awareness and Use

Reported factors that prevented portal use were lack of awareness, difficulty with setup, and lack of belief in portal usefulness, rather than lack of technological access. Patient education can address some barriers to patient portal use. However, lower perceived usefulness and lower confidence in personal use are more complicated barriers, which may be addressed through modification of patient portal designs to be as intuitive and simple as possible [24]. Tools such as the Centers for Disease Control and Infection’s Clear Communication Index can be used to identify the effectiveness of web-based health information and has been used to assess quality in patient portals and improve their simplicity and clarity [25]. Furthermore, eHL screening and in-person introductions to portals may improve portal uptake [26,27].

Study Limitations

Because of limited patient surveys administered before the onset of the COVID-19 pandemic, this study is underpowered to detect changes in portal uptake as a result of COVID-19. Other limitations of this study include relying on self-reported measures of technology access and past patient portal use, and that this study population did not include many individuals with technology access barriers. This single-site study represents patients of a large Midwestern, academic, urban medical center that may not be generalizable to suburban and rural patient populations in other regions or countries. Furthermore, generalizability to all inpatients may be limited as the sample was comprised primarily of adults hospitalized during the height of the COVID-19 pandemic. Finally, the use of eHEALS may fail to capture more dynamic, modern components of digital competency that newer scale measures such as the Digital Health Literacy Instrument, eHealth Literacy Questionnaire, and eHealth Literacy Assessment Toolkit [17,28].

Conclusions

In conclusion, this study indicates that low eHL was strongly associated with decreased patient portal awareness, use, and more negative portal attitudes among adult hospitalized patients. As health care professionals increasingly rely on patient portals, eHL should be accounted for to ensure patients with lower literacy are not disproportionately disadvantaged. Future studies should aim to understand how patient portal design and provider communication surrounding patient portals can be optimized for patients with low eHL. Further investigation of what interventions increase individuals’ eHL may better equip patients to take advantage of growing health care technologies, although additional work on also empowering patients to do so is also needed.

Acknowledgments

We would like to thank Mary Akel for her assistance with this project. This work is funded and was supported by the NHLBI K23 (HL118151 01), a Chicago Center for Diabetes Translation Research at the University of Chicago Pilot and Feasibility Grant, and a NORC/University of Chicago Center on Demography and Economics of Aging Pilot Award. The Hospitalist Project is funded and supported by Cultivating Health & Aging Researchers by Integrating Science, Medicine & Aging (5R25AG060910-04), the Center for Health Aging Behaviors and Longitudinal Investigations (5P30AG066619-03), and a Clinical and Translational Science Awards grant. VP reports receiving funding from the National Institutes of Health (R01HL146644) and the Agency for Health Care Research and Quality (R01HS027804).

Conflicts of Interest

VP reports consultant fees for Vizient Inc and Humana.

Multimedia Appendix 1

Summary of patient portal survey items.
[DOCX File, 16 KB - humanfactors_v10i1e40105_app1.docx]

Multimedia Appendix 2

Summary of eHEALS scale items.
[DOCX File, 17 KB - humanfactors_v10i1e40105_app2.docx]

References


Abbreviations

eHEALS: eHealth Literacy Scale
eHL: eHealth literacy
OR: odds ratio

©Nikita Deshpande, Vineet M Arora, Hanna Vollbrecht, David O Meltzer, Valerie Press. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 27.01.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Barriers to Telemedicine Use: Qualitative Analysis of Provider Perspectives During the COVID-19 Pandemic

Milan Patel1*, MD; Hanna Berlin2*, MD; Abishek Rajkumar2*, MD; Sarah L Krein3,4, RN, PhD; Rebecca Miller5; Jessie DeVito5, MS, MD; Jake Roy5, BS; Margaret Punch6, MD; Chad Ellimootti4,7, MS, MD; Alex F Peahl4, MSc, MD

1Department of Urology, Loyola University Medical Center, Maywood, IL, United States
2University of Michigan Medical School, Ann Arbor, MI, United States
3Center for Clinical Management Research, VA Ann Arbor Healthcare System, Ann Arbor, MI, United States
4Institute for Healthcare Policy and Innovation, University of Michigan, Ann Arbor, MI, United States
5Virtual Health, Michigan Medicine, Ann Arbor, MI, United States
6Department of Obstetrics and Gynecology, University of Michigan, Ann Arbor, MI, United States
7Department of Urology, University of Michigan, Ann Arbor, MI, United States

*these authors contributed equally

Abstract

Background: Though telemedicine is a promising approach for removing barriers to care and improving access for patients, telemedicine use for many medical specialties has decreased from its peak during the acute COVID-19 public health crisis. Understanding the barriers and facilitators to the maintenance of web-based visits—one key component of telemedicine—is critical for ensuring the continuous availability of this service for patients.

Objective: The purpose of this study is to describe medical providers’ perceived barriers and facilitators to the continued use of web-based visits to inform quality improvement efforts and promote sustainability.

Methods: We performed a qualitative content analysis of free-text responses from a survey of medical providers administered from February 5-14, 2021, at a large, midwestern academic institution, including all providers from medical professions that offered telemedicine (eg, physicians, residents or fellows, nurse practitioners, physicians assistants, or nurses) who completed at least 1 web-based visit from March 20, 2020, to February 14, 2021. The primary outcome was the experience of providing web-based visits, including barriers and facilitators to continued usage of web-based visits. Survey questions included 3 major domains: quality of care, technology, and satisfaction. Responses were coded using qualitative content analysis and further analyzed through a matrix analysis to understand the providers’ perspectives and elucidate key barriers and facilitators of web-based visit usage.

Results: Of 2692 eligible providers, 1040 (38.6%) completed the survey, of whom 702 were providers from medical professions that offered telemedicine. These providers spanned 7 health care professions and 47 clinical departments. The most common professions represented were physicians (486/702, 46.7%), residents or fellows (85/702, 8.2%), and nurse practitioners (81/702, 7.8%), while the most common clinical departments were internal medicine (69/702, 6.6%), psychiatry (69/702, 6.6%), and physical medicine and rehabilitation (67/702, 6.4%). The following 4 overarching categories of provider experience with web-based visits emerged: quality of care, patient rapport, visit flow, and equity. Though many providers saw web-based visits as a tool for improving care access, quality, and equity, others shared how appropriate selection of web-based visits, support (eg, patient training, home devices, and broadband access), and institutional and nationwide optimization (eg, relaxation of licensing requirements across state borders and reimbursement for phone-only modalities) were needed to sustain web-based visits.

https://humanfactors.jmir.org/2023/1/e39249
Conclusions: Our findings demonstrate key barriers to the maintenance of telemedicine services following the acute public health crisis. These findings can help prioritize the most impactful methods of sustaining and expanding telemedicine availability for patients who prefer this method of care delivery.

(JMIR Hum Factors 2023;10:e39249) doi:10.2196/39249

KEYWORDS

telehealth; virtual visits; public health crisis; barriers and facilitators; provider perspectives; implementation; access; telehealth; health care; patient care

Introduction

The COVID-19 pandemic has triggered remarkable growth in telemedicine. Telemedicine services such as web-based visits offered health care delivery alternatives that limited viral exposure and the use of resources while maintaining necessary medical services. Health care workers across professions and specialties rapidly adopted telemedicine to provide diverse services for their patients—from prenatal visits to preoperative consults to psychiatric counseling [1-3]. As a result, web-based visits increased from <1% of all outpatient encounters prior to the public health crisis to at least 30% in the initial months of the pandemic [4-7].

Telemedicine is a promising avenue for improving patients’ health care convenience and access by reducing care barriers like travel and childcare needs [8-12]. Yet, following its rapid uptake in the acute public health crisis, telemedicine usage has declined for some specialties, while others, like psychiatry, have maintained high levels of usage [6,7,13]. Clearly, some decline in telemedicine services was expected with the relaxation of social distancing and decreased risk of viral exposure. Specialties that do not require regular physical examinations or laboratory data may be more conducive to telemedicine; however, there are at least some applications for telemedicine in all specialties, from incision checks to medication adjustments. The uneven decline in the use of telemedicine suggests that other factors may contribute to whether practices maintain even modest levels of telemedicine offerings. The challenges at the patient, provider, and institution levels may preclude the maintenance of telemedicine services following the acute public health crisis. Specifically, concerns about care quality, supporting technology, and equity have been highlighted as the potential roadblocks to the continued widespread availability of telemedicine services [1,14,15]. Yet to date, the barriers and facilitators of sustained telemedicine usage are insufficiently described.

Health care professionals who deliver telemedicine services are uniquely positioned to understand the multilevel barriers, facilitators, and solutions needed to support their continued use. Our institution rapidly scaled web-based visits during the acute COVID-19 pandemic, increasing from 22 visits per day in February 2020 to a peak of 1823 per day in December 2021, with a subsequent decline. Thus, we conducted a mixed methods survey of providers to understand the drivers of telemedicine maintenance and inform quality improvement efforts necessary to support continued telemedicine services.

Methods

Ethical Considerations

This study was a qualitative content analysis of the free-text responses collected from a provider survey administered during February 5-14, 2021, to providers at Michigan Medicine who had completed at least 1 web-based visit. The University of Michigan Institutional Review Board deemed this survey project unregulated.

Overview

In response to the COVID-19 public health crisis, in March 2020, our institution rapidly scaled web-based visit capability for all providers (eg, physicians, nutritionists, and social workers). The providers were encouraged to use the recommended electronic health record–based platform; however, in concordance with emergency Health Insurance Portability and Accountability Act (HIPAA) regulations, other communication technologies (eg, Zoom) were deemed acceptable. A web-based help desk was available for technological issues. Individual specialties varied in web-based visit implementation, including (1) provider location (home or in office), (2) visit scheduling (web-based–only blocks, interspersed with in-person visits), (3) patient rooming (medical assistant [MA]–reviewed patient information, no MA check-in), and (4) providers’ use of clinic or personal devices.

This survey was developed with the Virtual Care Team at our institution for quality improvement. Questions addressed the key domains previously identified as the potential drivers of telemedicine maintenance, including the following:

1. Quality: providers’ ability to deliver medical services and develop rapport with patients
2. Technology: the degree to which patients and providers were able to use and complete video visits;
3. Satisfaction: providers’ overall experience with video visits and willingness to continue them following the acute public health crisis, including burnout and payment
4. Equity: the effect of telemedicine on existing health care inequities

Questions were asked in multiple-choice format, with free-text responses available. All questions were reviewed by an expert in telemedicine (CE) and survey methodology (AP). The survey was pilot-tested and approved by local telemedicine champions, with no recommended revisions, prior to deployment. The final survey included 7 multiple-choice questions and 2 free-text response questions. Participants were also able to provide free-text responses to give further context to their
multiple-choice selections (Multimedia Appendix 1). All qualitative responses were included in the content analysis. The survey was administered through the web-based Qualtrics platform.

We performed a qualitative content analysis of the free-text responses. Qualitative data were uploaded to MAXQDA software (version 20.4.0; VERBI GmbH) for management and analysis. Three authors (AP, MP, and HB) immersed themselves in the data and generated a preliminary codebook using inductive reasoning to construct initial codes using the constant comparison method [16-18]. This codebook was applied to the first 50 free-text responses for each question. The 3 authors then met to discuss the codebook, resolve discrepancies, and develop definitions and examples for each code. Following this initial coding consensus meeting, code definitions were revised and 3 codes were added to the codebook, for a total of 63 codes. Two authors (MP and HB) jointly coded groups of 20 responses until reaching 100% agreement, and then coded the remaining responses independently. The authors met frequently, and a third author (AP) resolved the coding discrepancies if necessary. The final codebook included 2 general groups of codes: codes describing individuals’ experiences with web-based visits and codes describing their ideal future state. For codes describing actual experiences, we used a matrix analysis technique to further understand the providers’ perspectives, comparing barriers and facilitators within each code in a grid. A comparison of these responses allowed us to identify the most salient drivers of telemedicine maintenance. The codes related to the providers’ ideal future state were presented separately as potential recommendations for improving telemedicine delivery.

**Results**

**Overview**

Of the 2692 providers at Michigan Medicine who had completed at least 1 web-based visit, 1040 (38.6%) completed the survey, including 702 providers from medical professions that offered telemedicine. These providers represented 4 health care professions, including physicians (486/702, 46.7%), residents or fellows (85/702, 8.2%), and nurse practitioners (81/702, 7.8%) from 47 clinical departments, including internal medicine (69/702, 6.6%), psychiatry (69/702, 6.6%), and physical medicine and rehabilitation (67/702, 6.4%; Table 1).

The following 4 overarching categories of provider experience emerged: quality of care, patient rapport, visit flow, and equity. See Figure 1 for key domains and Multimedia Appendix 2 for full quotations by domain.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Professions</strong></td>
<td></td>
</tr>
<tr>
<td>Physician</td>
<td>486 (46.7)</td>
</tr>
<tr>
<td>Resident or fellow</td>
<td>85 (8.2)</td>
</tr>
<tr>
<td>Nurse practitioner</td>
<td>81 (7.8)</td>
</tr>
<tr>
<td>Physician assistant</td>
<td>50 (4.8)</td>
</tr>
<tr>
<td><strong>Top clinical departments</strong></td>
<td></td>
</tr>
<tr>
<td>Internal medicine</td>
<td>69 (6.6)</td>
</tr>
<tr>
<td>Psychiatry</td>
<td>69 (6.6)</td>
</tr>
<tr>
<td>Physical medicine and rehabilitation</td>
<td>67 (6.4)</td>
</tr>
<tr>
<td>Neurology</td>
<td>51 (4.9)</td>
</tr>
<tr>
<td>Obstetrics and gynecology</td>
<td>50 (4.8)</td>
</tr>
<tr>
<td>Family medicine</td>
<td>39 (3.8)</td>
</tr>
<tr>
<td>General medicine</td>
<td>39 (3.8)</td>
</tr>
<tr>
<td>Hematology and oncology</td>
<td>38 (3.7)</td>
</tr>
<tr>
<td>Gastroenterology and hepatology</td>
<td>37 (3.6)</td>
</tr>
</tbody>
</table>

Table 1. Professions and top clinical departments of survey participants.
Quality of Care

Providers’ perceived ability to deliver high-quality medical care through telemedicine was mixed. While some believed “the quality of care is also same, if not better, than in person,” others found video visits were an inadequate substitution for in-person care, reporting, “the quality of video visits is garbage.” Perceptions of quality were driven by 4 key factors: appropriate selection of visit modality, availability of clinical services, patient population, and effects of web-based visits on communication.

For appropriately selected specialties (eg, psychiatry), visit types (eg, medication check-ins), and diagnoses (eg, chronic disease management), web-based visits could have the equal quality to in-person care. In some cases, appropriately selected web-based visits even improved patients’ care—seeing patients in their home environment provides a better examination than we get in clinics. This web-based “home visit” provided richer data than an office examination. Other times, web-based visits were an inadequate substitute for in-person care, such as for new patients or when specific laboratory data were needed.
Scheduling web-based visits when in-person care was more appropriate contributed to low-quality web-based visits for some patients: "...when it became clear a physical exam was necessary." More appropriate triaging of visits was seen as critical for "much more efficient experience[s]."

Providers highlighted physical examination maneuvers, tests, and additional services that could not be delivered digitally, including measuring limb strength, listening to fetal heart tones, and electrocardiograms. Web-based physical examinations had several shortcomings, including poor lighting, image quality, and camera angles. Similarly, providers found it difficult to get a "gestalt" of how a patient was doing through body language and nonverbal cues. Finally, providers noted coordination of services (eg, vaccines and referrals) was more complex with web-based visits: "I can’t walk down the hall and ask the social worker if she can see the pt while she is in clinic." Not being able to "walk down the hall" meant providers needed to spend more time "on the back end" arranging care for their patients.

Web-based visits were seen as more appropriate for some patient groups than others. For example, patients who were sick, with physical disabilities, were concerned about viral exposure or had out-of-town family members greatly benefited from web-based visits. In contrast, providers worried that other patients, including children and people with a history of hearing loss or trauma, might be disadvantaged by this modality.

Some providers noted improved communication in web-based visits, saying they allowed "more time to address their [patients'] issues." Others noted how web-based tools, including screen sharing, facilitated better patient counseling. One provider commented on how web-based counseling translated into improvements in health behaviors: "I have been successful in counseling, and when I see the patients back in office, am pleasantly surprised with improvements reported by changes made with behavioral modifications discussed during initial video visit."

Not all providers experienced improvements in patient counseling. Some noted how the web-based environment made teaching more difficult as visual aids and hands-on assistance with devices (eg, inhalers and insulin pumps) were less accessible. Additionally, nonclinical distractions, including driving, family members, or a lack of privacy, made conversations more difficult between patients and providers, leading providers to "worry that pieces of information may be getting lost." These patients could not fully engage in web-based conversations.

**Patient Rapport**

Web-based visits both improved and challenged providers’ ability to connect with their patients. Some noted equal or even improved patient rapport through web-based visits. These positive experiences were driven by the ability to see patients in their home environment, which "creates a new kind of intimacy, kind of like a modern-day house call." The web-based "house call" provided new depth to the patient-provider relationship beyond the sterile clinical setting. Other providers noted patients seemed more at ease in their own homes "rather than being in an office setting as a ‘patient.’" Being able to see patients’ faces and better interpret their facial expressions was also advantageous during the pandemic mask mandates.

In contrast, some providers found it challenging to connect in the web-based environment, reflecting that "this technology is eroding the doctor-patient relationship significantly." Providers identified 3 sources of this challenge: lack of privacy, invasiveness of web-based visits, and the inability to connect through touch. Providers reported that patients sometimes could not find a private space for their visit or were multitasking, making it particularly difficult to have a safe space to build a patient-provider relationship. Similarly, the invasion of privacy in web-based visits was intrusive for some patients: "Many people struggle with clutter and are embarrassed to have people in the home." Thus, for some patients, web-based visits represented a loss of the "neutral space" and subsequent comfort provided by clinical settings.

Providers also commented on the loss of "laying on of hands," which they saw as an important part of the therapeutic alliance. This was particularly true for providers in specialties where the physical examination was central to the visit: "I am in a specialty that requires more hands-on physical exam. That helps build trust and rapport that is impossible over video." As a result, building equal trust and rapport in a web-based environment felt impossible for some providers.

**Visit Flow**

Providers reported that web-based visits improved efficiency through better patient show rates, eliminating patient care delays, and allowing for real-time documentation. The ability to easily connect to technology was crucial for realizing these gains. Additionally, internal resources, including MA rooming assistance for some visits or a medical student initiative that provides support to geriatric patients connecting to web-based services, were considered helpful for ensuring patients could complete web-based visits.

Some providers noted reduced appointment cancelations and patient no-shows for their web-based visits compared to in-person care, particularly for patients with barriers to care like low-income patients, those with disabilities, and those who lived far from the clinic. Web-based visits also alleviated other common reasons for missed appointments, like weather or traffic delays.

Providers noted other advantages to appointment efficiency, including time saved rooming patients: "There is less transition time, and I'm spending more time with their patients as opposed to going from point A to point B." Less "transition time" left providers with more time with their patients. For some providers, web-based visits also facilitated more efficient documentation, as they could maintain eye contact while typing. Improved efficiency contributed to higher satisfaction for both providers and patients. Providers noted how reduced travel time and coordination led to a better patient experience.

*If it is a counseling or other appointment that an exam is not required, I would argue that the care is exactly the same if not better as patient's don't have the added stress of travel/parking/checking in/out.*
The combination of improved efficiency and higher patient satisfaction led to improved job satisfaction for many providers: “This leads to substantially higher satisfaction for patients and myself... which improves my home life as well.” In sum, for many providers, web-based visits offered a pathway to more efficient, streamlined patient care, resulting in a better patient and provider experience.

Some providers experienced frustrating challenges with web-based care, including higher no-show rates, technical difficulties, and insufficient support—all resulting in lower satisfaction. The providers shared a variety of possible reasons for higher no-show rates for web-based visits, including administrative errors, technical difficulties, or patients’ perceived differences between canceling in-person and web-based visits.

Additionally, providers perceived delays in care due to insufficient patient and provider training on web-based platforms and also expressed issues accessing the electronic health record app, using the software, and even turning on video and sound. These burdens were felt even more when web-based visits were interspersed with in-person care. Similarly, providers described a need for better real-time technical support for patients “when they are struggling” at the beginning of a visit. One provider suggested the use of a “tech-barrier interpreter,” similar to interpreters used for patients with language barriers.

Web-based visits were made even more “hectic” by the lack of support some providers perceived in reviewing the patient’s history and medications, completing questionnaires, and gathering historical data. Though these processes were completed through an automated e-check-in, providers perceived this information as less reliable than in-person data entered by MAs, clerks, and trainees, who provided at-the-elbow assistance for in-person care.

Other provider frustrations with clinical workflow included more onerous documentation, technical difficulties, and an increased volume of postvisit follow-up work. Some providers noted challenges in documenting in real time through web-based visits and reported more unresolved postvisit documentation. Providers noted issues with inconsistent internet connections and the need to convert to telephone encounters from video visits, resulting in “frequent technical hurdles which are frustrating to patients and providers.” As a result, some providers saw web-based visits as inefficient and disruptive.

Equity

Providers saw web-based visits as both a facilitator and a barrier to equity. Providers highlighted how web-based visits improved access for specific populations, including patients with disabilities, rural patients, and patients with poor access to care. Still, some providers worried that web-based visits might further exacerbate existing inequities for patients without access to devices or broadband or who did not speak English. As a result, some providers worried that the push for web-based care would result in some patients being left behind. As one provider reflected, “I do worry a lot about my patients who have limited technology access.”

Many providers emphasized the need to maintain video and phone visit options to ensure access for all patients, noting specific populations that might benefit from phone visits, including older adults, patients with disabilities or low technology literacy, and patients without access to needed technology (eg, those in rural areas).

Finally, providers shared that some populations, particularly those who have faced historical injustices, expressed concerns about privacy. As one described, “there is also some mistrust of VV [virtual visits] in our at-risk populations. I have had some tell me they don’t want the VV option, just phone, as they are concerned about being recorded.” This fear of inadequate privacy led some patients who may have benefited from web-based visits to decline the service. Rather than improving convenience and access, web-based visits “eroded the doctor-patient relationship significantly,” increasing existing inequities.

Ideal Future State

Providers noted several institutional and national changes needed to realize the ideal future state of telemedicine delivery (Table 2). Institutionally, providers emphasized the importance of providing patients with tools to ensure web-based visits were of the highest quality, including a loaner program for tablets, having home visiting nurses complete vital signs and components of the physical examination, and helping patients obtain broadband access. Some providers perceived that institutional targets for web-based visits challenged providers’ autonomy to determine visit type, limited the patient-centeredness of care, and at worst, reduced patients’ access to appropriate visit types: “Having arbitrary goals for the number of virtual visits is insulting. It suggests that we should practice to achieve a metric.” From the provider’s perspective, incentivizing “arbitrary goals” for web-based care limited their ability to use shared decision-making to decide on appropriate web-based visit use.

On the state and national levels, providers identified 3 key policy recommendations: (1) removal of state-based licensing restrictions, (2) maintenance of parity for video- and audio-only visits, and (3) support for broadband expansion and access. Providers noted how state-based licensing prevented patients who were “traveling the farthest” from taking advantage of web-based visits. This limited the continuity of care for the patients who traveled during the winter or lived over state lines. As emphasized above, parity between video- and audio-only visits was a top priority for maintaining equity in new models. Finally, providers shared ideas for partnering with lawmakers and communities to expand broadband access. While some highlighted the need to “invest in universal access to broadband internet access, just like water and electricity and roads,” others envisioned building infrastructure in community physicians’ offices or libraries. In sum, providers envisioned a future where policy supports equitable access to high-quality web-based visits for all patients who prefer this modality, across state borders, and through both video- and audio-only platforms.

https://humanfactors.jmir.org/2023/1/e39249
Table 2. Specific policy concerns.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Concern</th>
<th>Ideal future state reflected by providers</th>
</tr>
</thead>
<tbody>
<tr>
<td>No out-of-state limits</td>
<td>Providers noted how licensing barriers prevented accessibility for out-of-state patients.</td>
<td>• Difficult for hospital with large catchment area like U of M to not be able to offer video visits for people traveling the farthest.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The other group is patients who live out of state—enabling providers to see these patients would improve access to and continuity of care (for example when patients go to Florida for the winter, or those who live in Toledo).</td>
</tr>
<tr>
<td>Parity of audio or video</td>
<td>Maintaining equity for video- and audio-only visits was seen as a critical issue for ensuring access and equity.</td>
<td>• Lobby for phone reimbursement equivalent to other forms of care, since telephone is often more accessible than video connection.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The obsession with video visits definitely is a DEI issue. The more disadvantaged of our patients are the ones less likely to be able to do video visits and I am shocked the University has not been more aware of this. A good way to help this would be to endorse phone visits. Everyone has and knows how to use a telephone. There are seldom technical issues when calling someone on a phone. Stop disparaging phone visits.</td>
</tr>
<tr>
<td>Need for public partnership for broadband</td>
<td>Many providers saw policy changes and partnerships with community institutions as promising avenues for expanding access to video visits.</td>
<td>• Encourage our legislature to invest in universal access to broadband internet access, just like water and electricity and roads. The internet is a necessary utility and not a luxury, and pandemic should have erased any doubt about this.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• For rural communities that have poor or no internet, would it be worth providing pts with a list of sites where they could find a private room with internet (e.g., library)? Similarly, for low income communities (or maybe ALL communities), would it be worth providing a list of sites where they could find a private room with computer &amp; internet (e.g., library)?</td>
</tr>
<tr>
<td>Need for patient education</td>
<td>Providers identified a need for classes to help patients with technology skills.</td>
<td>• Would be nice to offer a virtual video visit support (101) class for those who want to become more tech savvy.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• If we could continue to optimize patient education resources or real-time assistance to help them with establishing the video connections, that would be great, and then it would be 100% of the way there. I’d say it’s 80%-90% of the way there right now. :)</td>
</tr>
<tr>
<td>Need for devices</td>
<td>Providers saw devices as crucial for improving patients’ access to video visits.</td>
<td>• Providing patients needing frequent video visits with devices (loaners). Facilitating internet access for families.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Consider pushing out technology (low-cost tablets with 4G or 5G capability) to our patients in high poverty areas.</td>
</tr>
<tr>
<td>Flexible institutional policy</td>
<td>Providers noted frustration with several institutional video visit policies, including establishing a required proportion of video visits and removing needed social support services.</td>
<td>• Michigan Medicine has also unfortunately reduced prioritization of social work support in the health system and have cut staff who previously had the time to take the extra time to work with patients and families who had less access and lower resources but now they are required to move too quickly through their work in scheduling to even learn of people’s needs. It costs money to serve underserved populations and unless the institution backs this priority with financial resources, this will be words and not action.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• I strongly feel that the number of virtual visits should be determined by shared physician or provider and patient decision-making. Having arbitrary goals for the number of virtual visits is insulting. It suggests that we should practice to achieve a metric rather than practice in a way that is medically reasonable and personally acceptable to patients.</td>
</tr>
</tbody>
</table>

Discussion

Principal Results

In this cross-sectional survey, providers across multiple specialties and professions shared the benefits and challenges of providing telemedicine in 4 key domains: quality, patient rapport, visit flow, and equity. These nuanced experiences highlight how telemedicine may help address many pressing issues in health care but is not a stand-alone solution. As we move further away from the COVID-19 public health crisis, how to best integrate telemedicine, and to what extent and for which specialties, remains unknown [19,20].

Comparisons With Other Works

Maintaining high-quality health care is of utmost importance. In our study, many providers reported that web-based visits can be a high-quality delivery method for specific appointment types (eg, return visits and medication check-ins) and specific conditions (eg, chronic disease management and mental health needs). For some providers, web-based visits actually improved the quality of care by allowing for more real-world counseling and advice in the patient’s home environment. However, for visits where additional data are required, providers echoed concerns previously raised in the literature that web-based visits may provide a substandard level of care [21-23]. A deeper understanding of what visits are appropriate for telemedicine and better triaging of visit types are crucial steps to ensuring the quality of care is maintained. Similarly, novel approaches to making components of in-person visits available at home through provider training, making home devices available, and expanding options for laboratory testing and imaging may be important methods to improve the availability of objective data for web-based visits [24-27].
Providers have previously expressed concerns about building patient rapport through web-based visits [21]. While some providers in our study echoed these concerns, others noted how web-based visits helped facilitate patient-provider connections through improved patient comfort and the ability to create greater intimacy with a web-based “house call.” Rapport depended on patients having a safe, private space for web-based visits, which was less possible in specialties where the physical examination was central to connection. Incorporating patients’ perspectives will be critical for further understanding relationship-building in the web-based environment and what, if any, changes in health behaviors and outcomes result.

While telemedicine promises to alleviate inefficiencies and inequities by reducing travel and other barriers to in-person care delivery [9,11,12,28], many providers in this study shared how lagging technology, insufficient technical and clinical support, poor connectivity, and digital literacy issues prevent this ideal from being realized. It is clear that to maintain telemedicine services, greater web-based infrastructure is needed, including higher-quality internet connections, more robust training, and ensuring services that in-person staff provide are also accounted for digitally. Similarly, the digital divide threatens to worsen, rather than alleviate, care disparities if efforts are not made to ensure patients at greatest risk of adverse outcomes are not left behind [28].

Providers highlighted several critical local and national policies to maintain telemedicine beyond the acute public health crisis. Some solutions, including more flexible messaging around organizational targets for web-based care adoption and maintaining parity for audio-only and video encounters, highlight the need for tailored care that is responsive to patients’ preferences. Other solutions, including public partnerships to ensure access to broadband, highlight the stark inequities in our current health care system. Several of the policy priorities emphasized in our study have been considered in local and national discussions; however, few protections exist for telemedicine gains made during the public health crisis without major legislative change at the state and federal levels [19,20].

**Limitations**

Our study has several limitations. It was conducted at a single institution and may not reflect the perspectives of providers in other geographic locations, serving different patient populations, or in settings with more or less robust telemedicine infrastructure. Additionally, providers who had particularly strong opinions about telemedicine may have been more likely to complete our survey, creating selection bias. Still, we believe that the volume of our qualitative data from diverse experiences provides rich insights to inform the maintenance of telemedicine. Another limitation of our study is that we did not investigate patient perspectives. Future studies should focus on comparing the opinions of both patients and providers to capture the total user experience.

**Conclusions**

Web-based visits provide an important opportunity to improve care quality, connection, efficiency, and equity. However, significant challenges threaten to erase gains made in the provision of telemedicine during the public health crisis, particularly for specialties that require some in-person services. Our study highlights providers’ perceptions of the most important local and nationwide efforts needed to maintain web-based visits beyond the COVID-19 pandemic. Through these adaptations, health care can meet patients where they are with high-quality, equitable, and patient-centered services.

---

**Acknowledgments**

CE was supported by the Agency for Healthcare Research and Quality (AHRQ; grant #K08 HS027632). The AHRQ played no role in the conception or design of the study, the collection, management, and analysis of the data, the preparation or review of the manuscript, or the decision to submit the manuscript for publication.

**Conflicts of Interest**

AFP is a paid consultant for Maven Clinic. The remaining authors declare no conflicts of interest.

---

**Multimedia Appendix 1**

Provider survey.

[DOCX File, 26 KB - humanfactors_v10i1e39249_app1.docx ]

**Multimedia Appendix 2**

Matrix coding of provider telemedicine experience.

[DOCX File, 86 KB - humanfactors_v10i1e39249_app2.docx ]

---

**References**


Investigating the Connections Between Delivery of Care, Reablement, Workload, and Organizational Factors in Home Care Services: Mixed Methods Study

Adam S Darwich¹, MSc, PhD; Anne-Marie Boström²,³,⁴, RN, MSc, PhD; Susanne Guidetti⁵,⁶, MSc, PhD; Jayanth Raghothama¹, MTech, PhD; Sebastiaan Meijer¹, MSc, PhD

¹Division of Health Informatics and Logistics, Department of Biomedical Engineering and Health Systems, KTH Royal Institute of Technology, Huddinge, Stockholm, Sweden
²Division of Nursing, Department of Neurobiology, Care Science and Society, Karolinska Institutet, Huddinge, Sweden
³Theme Inflammation and Aging, Karolinska University Hospital, Huddinge, Sweden
⁴Research and Development Unit, Stockholms Sjukhem, Stockholm, Sweden
⁵Division of Occupational Health, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden
⁶Theme Women's Health and Allied Professionals, Medical Unit Occupational Therapy and Physiotherapy, Karolinska University Hospital, Stockholm, Sweden

Corresponding Author:
Adam S Darwich, MSc, PhD
Division of Health Informatics and Logistics
Department of Biomedical Engineering and Health Systems
KTH Royal Institute of Technology
Hälsvägen 11C
Huddinge, Stockholm, 141 57
Sweden
Phone: 46 8 790 48 05
Email: darwich@kth.se

Abstract

Background: Home care is facing increasing demand due to an aging population. Several challenges have been identified in the provision of home care, such as the need for support and tailoring support to individual needs. Goal-oriented interventions, such as reablement, may provide a solution to some of these challenges. The reablement approach targets adaptation to disease and relearning of everyday life skills and has been found to improve health-related quality of life while reducing service use.

Objective: The objective of this study is to characterize home care system variables (elements) and their relationships (connections) relevant to home care staff workload, home care user needs and satisfaction, and the reablement approach. This is to examine the effects of improvement and interventions, such as the person-centered reablement approach, on the delivery of home care services, workload, work-related stress, home care user experience, and other organizational factors. The main focus was on Swedish home care and tax-funded universal welfare systems.

Methods: The study used a mixed methods approach where a causal loop diagram was developed grounded in participatory methods with academic health care science research experts in nursing, occupational therapy, aging, and the reablement approach. The approach was supplemented with theoretical models and the scientific literature. The developed model was verified by the same group of experts and empirical evidence. Finally, the model was analyzed qualitatively and through simulation methods.

Results: The final causal loop diagram included elements and connections across the categories: stress, home care staff, home care user, organization, social support network of the home care user, and societal level. The model was able to qualitatively describe observed intervention outcomes from the literature. The analysis suggested elements to target for improvement and the potential impact of relevant studied interventions. For example, the elements “workload” and “distress” were important determinants of home care staff health, provision, and quality of care.

Conclusions: The developed model may be of value for informing hypothesis formulation, study design, and discourse within the context of improvement in home care. Further work will include a broader group of stakeholders to reduce the risk of bias. Translation into a quantitative model will be explored.
Introduction

The home environment is the preferred care setting for many older adults [1]. Home care services can be offered in several instances without compromising health outcomes at a significantly lower cost as compared to institutional care [2]. Health care systems and associated payer-models differ between countries [3]. In Sweden, home care for older adults forms a part of the universal welfare system regulated on a national level by the Swedish Social Services Act (SFS 2001:453) [4,5]. Services are provided on an individual need basis and should offer high-quality care to all citizens aged 65 years and older. In practice, regulations are enacted on a local level in the regions and municipalities responsible for financing and governing home care services provided by public, nonprofit organizations, and private companies. The quality and extent of offered home care services may therefore vary across the country based on regional finances and policy [6,7].

As the overall life expectancy increases and the population distribution shifts toward older age, more pressure is being put on already resource-constrained health care systems and welfare services. Treatment outcomes are improving, meaning that the type of care needed is shifting toward management of chronic disease. Further, health care systems are moving toward distributed care models and treatment at home, enabled by technological innovations [5,8]. In Sweden, home care services are subject to increased demand and growing complexity of responsibilities and tasks. Home care staff are faced with caring for an increasing number of home care users, leading to higher workload, stress, and burnout [9].

Home care services are adapting in part by adopting goal-oriented interventions [10]. Reablement is a person-centered approach to enhance an individual’s physical and other functioning, to increase or maintain their independence in meaningful activities of daily living at their place of residence and to reduce their need for long-term services [11]. The approach has been shown to improve health-related quality of life while reducing service use [12]. Reablement is an inclusive approach irrespective of age, capacity, diagnosis, or setting, and has been used for different population groups, including a growing field of reablement for people with dementia [11,13,14]. However, there is a need to evaluate the outcomes and effects of reablement to determine its benefit in specific population groups.

Predicting the impact of change in health services is challenging. Health care can be viewed as a complex adaptive system, with intricate relationships between individual variables of the system, feedback loops, and emergent behavior. Therefore, the design of new interventions, policies, and improvement benefit from a systems perspective. Systems thinking techniques, such as causal loop diagrams (CLDs), provide a framework for studying these systems, including context and how they respond to change. The approach usually involves mapping of constituent parts (referred to as “elements”) and their causal relationships (or “connections”), analyzing feedback loops, and using simulation techniques to investigate system behavior. Reinforcing feedback loops are potential targets for policy change due to their properties as leverage points. Model development can be carried out through detailing the current body of evidence, using documentation and other knowledge bases, participatory methods with domain experts, or a combination of these. As such, the approach is useful for integrating “hard” and “soft” knowledge into the decision-making process [15-22].

This work is part of the Future Care research program, aiming to contribute to the development of knowledge-based care, participation, and social inclusion for older adults. This includes studies on the working environment in home care, reablement with the support of information and communication technology (ICT; the ASSIST project), social participation, the design of physical spaces, and more [23,24].

The aim of this study is to characterize the home care systems elements and their connections relevant to home care staff workload, work-related stress, home care user needs and satisfaction, and the use of enabling technologies. This is to provide a systems model for examining the effects of interventions in home care, including the reablement approach and accompanying ICT (ASSIST, Future Care) [24], disabled home care users, home care staff, and organizational factors. As such, the project intended to take a broad, holistic perspective to improvement in the home care setting. A CLD was developed grounded in expert knowledge and validated instruments. The developed model was then verified and analyzed.

Methods

Overview

Various approaches have been used for participatory model development [21,25]. This study used an iterative approach, combining group model building and targeted data collection (for additional details, see section S1 in Multimedia Appendix 1 [12,26-49]). In total, 11 experts participated across the activities. The experts were all active academic health care science researchers in nursing and occupational therapy, focusing on aging, home care, nursing homes, health services for older adults, social participation, public health, evidence-based care and the reablement approach. The participants did not include home care staff, older people in home care or their relatives. No personal data were collected, accessed, or analyzed for this study. Experts were engaged in their work setting and professional role to collect their views.
Participation was voluntary and could be terminated at any time. No track records between individuals and their statements have been kept beyond the initial data collection. Model development was preceded by input from the experts to define the requirements, scope and boundaries, key scientific literature, and documentation to support model development and verification. This was followed by collection of literature data, model development, and verification against intervention data sets. Expert review was carried out at 2 separate occasions, this to assess the model structure, literature sources, and verification results. Following each expert review phase, additional targeted literature searches were carried out along with model refinement. Finally, model analysis was carried out by analyzing feedback loops, the model behavior using social network analysis and simulations (Figure 1).

**Figure 1.** Workflow of the iterative model development process.

**Ethical Considerations**
The research only involved the opinions of professionals in a professional setting, so no ethical approval was necessary according to the Swedish Ethical Review Act through the Swedish Ethical Review Authority [50].

**Requirements, Scope, and Boundaries**
The model requirements and scope were defined together with the domain experts. It was decided that the model should be able to describe home care staff workload, work-related stress, provision of care and services, home care user needs, and satisfaction. Relevant elements included those linked to the home care organization, home care staff, and users. The purposes of the model were to enable further analysis of the data being generated within the research program and to allow the study of interventions related to the reablement approach and associated technologies.

Key literature on the reablement approach and workload or job strain, including the questionnaires and theoretical models, QPSNordic (the General Nordic Questionnaire for Psychological and Social Factors at Work), and SDCS (Strain in Dementia Care Scale) [51,52] were identified a priori by the experts. QPSNordic describes items related to social and psychological factors in the workplace, including leadership, organization, control and demand, social climate, role conflict, and more. The survey is used to investigate work conditions and health and support organizational change. The SDCS describes staff strain in residential dementia care; this includes variables and their impact on job strain, including balancing and competing needs, frustrated empathy, understanding and interpreting, emotional involvement, and recognition. SDCS was designed to aid the identification and study of interventions to improve staff well-being in residential care, among others. The survey tool has also been applied in the home care setting [53]. Preliminary literature searches were carried out to identify suitable models of work-related stress for the home care setting, including the stress of conscience and quality of care.

**Model Development**
Data collection was carried out using MEDLINE and PubMed. This is to identify quantitative and qualitative predictors of stress in home care, residential care, care for older people, dementia care, nursing homes, and related settings (for more information, see Multimedia Appendix 1). A CLD was developed in Kumu (Kumu Inc). The final model can be viewed as an interactive map [54]. The data set with the full reference list (Table S4 in Multimedia Appendix 2 [20,52,54-110]), the model export file and analysis (Tables S5 and S6 in Multimedia Appendix 3), and the simulation script (Matlab script file in Multimedia Appendix 4) are also available as supplementary material. The stress model was developed based on theoretical models of work-related stress derived from the preliminary literature search and discussions with experts [51,111]. Additional literature data were incorporated into the model by reviewing the structure of the model and adding new elements and relationships one at a time. This was carried out in several iterations to ensure consistency between data and CLD. Here, the inclusion of studies from nursing homes and residential care was justified as important supplemental data in the absence of evidence from the home care setting. An expert review was carried out to ensure relevance to home care (section S4 and Figure S5 in Multimedia Appendix 1). Model elements were grouped into categories. The final categories included organization, home care staff, home care users, stress, social support network (of the home care user), and societal level (Table 1). The basis of this categorization was the conceptual level of the individual elements, as per Dallner and colleagues [111].
Table 1. Categories of the causal loop model along with their description.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stress</td>
<td>The core stress model, describing stress response as a combined effect of job demand and control.</td>
</tr>
<tr>
<td>Home care staff</td>
<td>Describes the delivery of care and services, professional competence and experience, and direct interactions with the home care organization. In addition, this includes elements of the home care staff’s private life and the impact of stress on mental and physical health.</td>
</tr>
<tr>
<td>Home care user</td>
<td>Describes home care user needs, experience and expectations, physiological and mental health, and direct interactions with home care staff and social support network.</td>
</tr>
<tr>
<td>Organization</td>
<td>Includes leadership and organizational elements, home care strategies for delivering care and services, work planning and scheduling, and direct interactions with home care staff.</td>
</tr>
<tr>
<td>Social support network (of the home care user)</td>
<td>Describes the home care user’s social network, their experience, and interactions with the home care user, and how this influences informal care.</td>
</tr>
<tr>
<td>Societal level</td>
<td>This includes higher-level elements that are extrinsic to the home care organization, home care staff, and user. This category includes the impact of regional unemployment, income, and care capacity on home care, stigmatization toward the profession, and the effect of home care on health care spending.</td>
</tr>
</tbody>
</table>

Model Evaluation and Refinement

Model evaluation and refinement were carried out on 2 occasions. This consisted of verification of the model structure based on the identified studies of interventions and outcomes, expert input through participatory workshops, and model refinement based on feedback.

Verification was carried out based on identified literature on relevant clinical intervention outcomes in nursing homes and home care (including the reablement approach). This was carried out by identifying and qualitatively comparing scenarios and outcomes with model elements representative of the intervention along with the cascade reaction produced in the model and its ability to recover the observed outcomes.

Model Simulation and Analysis

Social network analysis was carried out [55]. Methods and results of the social network analysis are detailed in Multimedia Appendix 1. A simplified simulation algorithm based on Boodagian and colleagues [112], was implemented in Matlab (release 2022a; Mathworks). Key elements of interest (relevant to the study aims) were investigated for their ability to influence the model as a whole; these included “needs met,” “provision of care and services,” “workload,” “distress,” “person-centered care” (ie, the reablement approach) and “home-care-staff user adoption of home care technology” (ie, ICT to support the reablement approach). The full set of simulation results is detailed in section S6 in Multimedia Appendix 1.

Results

Overview

The review of the literature identified 914 nonunique relationships between relevant variables in home care for older people, based on 59 publications (Table S4 in Multimedia Appendix 2). Additional knowledge was considered through participatory model development with academic experts. The final model included 122 elements and 223 connections divided across six categories (defined in Table 1): (1) stress (elements: n=6), (2) home care staff (n=44), (3) home care user (n=28), (4) organization (n=26), (5) social support of the home care user (n=13), and (6) societal level (n=5; Tables S5 and S6 in Multimedia Appendix 3). Figure 2 shows the full CLD.
Stress

Several theoretical models of work-related stress were identified in the literature [113-116]. One of the most well-established theoretical models chosen to describe work-related stress here is the demand-control model developed by Karasek [117].

According to demand-control theory, job strain and negative stress (distress) will increase when the impact of demand outweighs control. Further, an interaction effect has been observed between control and demand on stress, where an increase in control outweighs the effect of demand [118]. This was implemented by allowing control and demand to affect both elements’ underload and overload.

It was agreed in the expert group that a model of job strain and stress should account for a nonmonotonous (U-shaped) relationship between stress and demand and control. This means that when demand is higher than control, distress occurs through work overload. When control outweighs demand boredom, induced by work underload, becomes a potential source of distress (underload; Figure 3). Multiple home care staff elements linked to the stress model. For example, several types of demand affected “job demand.” While the “ability to cope” was affected by “mastery of work and professional competence,” “job involvement,” “job satisfaction,” and through feedback, by the stress response. “Recovery” indirectly affected the “ability to cope” positively. The element was reliant on home care staff, “social support,” “personal life demands,” “shift work,” and “overtime and unscheduled work.” Additional information is given in section S2 in Multimedia Appendix 1.
Home Care Staff

The home care staff category (Figure 4) included elements of the provision of care and services, primarily based on the quality satisfaction model developed by Samuelsson and Wister [56]. “Job demand” was dependent on 3 elements: learning, decision, and quantitative demands according to QPSNordic [111]. These depended on several elements, including “complexity of work tasks,” “workload,” and “work pace.”

Several elements of the home care staff category were connected to the “ability to cope” and subsequently “control” in the stress category. This included elements such as “job satisfaction and involvement,” “mastery of work,” “home care staff health and safety risk,” and “stress of conscience.”

Physiological and mental health responses to “distress” and their impact on home care staff burnout were described [58,59]. In this feedback loop “work-life balance” and “recovery” played an important role in balancing “distress” [58].

Multiple organizational elements are linked to the home care staff category. For instance, “role conflict” and “recognition” both affected “job satisfaction” [111]; “deskilling” affected “complexity of work tasks,” “training and specialization,” “ability to cope,” and “distress;” as well as “home care user trust in home care staff” [58,60-62]. Organizational elements related to working with technology (“working with sophisticated technology” and “equipment problems”) were linked to the home care staff category and were associated with an increase in “complexity of work tasks” and “workload” [60].

“Provision of care and services” was split into continuity, suitability, availability, influence, and personal relation as determining elements of the delivery of care and services in home care, quality, and home care user satisfaction with care. An increase in the home care user element “needs met” led to a reduction in “delivery of care and services,” and home care staff “stress of conscience” in 2 balancing feedback loops. In addition, the home care user’s ability to communicate, the staff “mastery of work and professional competence,” and “language proficiency” had a positive influence on “confirming home care staff-user relationship and family communication.” This element was an important determinant of “home care user influence” and “trust in home care staff” [119].

Home Care User

The home care user category included individual needs, satisfaction with care, functional ability, and autonomy (Figure 5). Home care user “needs met” was determined by several elements related to instrumental and noninstrumental activities of daily living, social, emotional, information, treatment-related, and self-sufficiency needs. This was mainly based on an analysis of home care user needs by Keeling [63]. Needs met influenced home care user satisfaction with care, self-perceived health, and quality of life [64].
Figure 5. Elements and connections of the categories home care user and social support network (color-coded orange and green, respectively) of the final causal loop diagram. Arrows indicate the directional connections between elements. Positive and negative connections are displayed in solid arrows (+) and dashed lines (–), respectively.

A notable behavior of the home care user category related to the influence of “needs met” on adverse events or progression of the disease, where inadequate care led to an increase in the need for care in a reinforcing feedback loop. The emergence of “adverse events or progression of disease” due to unmet needs and “time spent in care” would then lead to a higher “likelihood of institutional care.” This would in turn affect “health care spending” in the societal level category.

Organization

Many of the elements and connections of the category organization were based on the QPSNordic model [111]. This category (Figure 2) covered elements, such as the impact of leadership, social climate, and role clarity on workload, job involvement, and satisfaction. For instance, “quality of supervision” had a positive effect on “empowering leadership,” “human resource primacy,” and “commitment to the organization.” This in turn leads to greater “recognition,” home care staff “job satisfaction,” and “involvement.”

This category also described the impact of “organizational slack” and organizational “emphasis on cost-effectiveness” on the overall organization and working conditions [60,65-70,119]. For example, “emphasis on cost-effectiveness” led to “inadequate income” and “deskilling” of home care staff, which in turn reduced “job satisfaction,” “training and specialization,” as well as “mastery of work and professional competence,” hence having a negative impact on the “provision of care and services.”

Social Support

The social support network of the home care user (Figure 5) influenced the “delivery of informal care,” therefore increasing home care user’s “needs met.” This was dependent on the “social support network of the home care user” and the “quality of life of the informal caregiver” [63,71]. Further, “needs met” reduced the “informal caregiver stress” and home care staff’s “stress of conscience” in addition to lowering the demand for the “provision of care and services” [63].
Therefore, if home care user “needs met” were fulfilled, this led to improved “quality of life of an informal caregiver,” enabling “informal care of home care user” in a reinforcing feedback loop.

**Societal Level**

Figure 2 includes elements of the category societal level. The home care user element “likelihood of institutional care” affected the societal level element “health care spending” positively. Other elements in this category are related to the labor market and turnover intention. This included county-level unemployment, care capacity, capital income, and “stigmatization toward profession.” These elements all defined “job turnover” in the home care staff category [60,72].

**Qualitative Verification**

Here is presented the verification against observed data from intervention studies related to organizational change in the nursing home setting (section S4 and Table S2 in Multimedia Appendix 1) and reablement in home care (section S4 and Table S3 in Multimedia Appendix 1). Scenarios from the peer-reviewed literature were compared with causal pathways of the CLD to assess agreement.

The final CLD largely supported the intervention-outcome combinations of the observed studies on a qualitative level. The final model showed improvement regarding consistency with the observed studies as compared to earlier iterations. No changes were seen at the final iterative stage. For example, Burgio and coworkers [26] studied the impact of nursing staff training for general communication skills, a motivational system, recognition of staff, and feedback on communication skills. This led to improved staff communication skills, positive staff statements, and a higher degree of independent self-care among residents. In the initial CLD, “training and specialization” led to improved “home care user and family communication,” and reduced “social needs” of the home care user. Also, an increase in “support from superior” led to staff “job satisfaction.” In the final CLD, “training and specialization” increased “confirming home care staff-user and family communication and relationship,” increasing the “functional ability and autonomy” of the home care user. Then, “support from superior” improved staff “job satisfaction.”

Most notable were the improvements in the model’s ability to capture intervention outcomes for the reablement approach (model element, “person-centered approach”). For example, Burton and colleagues [30] observed an improvement in home care user health-related quality of life and a reduction in home care needs following the introduction of a reablement approach [30]. The first version of the CLD was able to recover the impact of “functional ability” on “delivery of care.” The final model could describe the impact of the “person-centered approach” on the “provision of care and services” and home care users’ “quality of life.” The full list of comparator studies and related model pathways are given in Tables S2 and S3 in Multimedia Appendix 1.

**Simulations**

A simulation analysis was performed to examine the theoretical impact of key elements of interest on the model (section S6 in Multimedia Appendix 1). The activation of “person-centered care” (Figure S7 in Multimedia Appendix 5) led to the activation of several elements related to delivery and quality of care and services, including “trust in home care staff,” “provision of care and services,” home care user “needs met,” and more. In addition, this led to a reduction in “home care staff health and safety risks.” On the other hand, the activation of the element “workload” led to a reduction in “provision of care and services,” “needs met,” and “job satisfaction.” Elements related to work demand, such as “quantitative demands,” “decision demands,” “work pace,” and “role conflict” increased. “Distress” remained unmodulated (see Discussion section). Simulating an activation of “home care staff-user adoption of technology” positively impacted factors related to “complexity of work tasks” along with “person-centered care.” Activation of “distress” led to a drop in elements related to home care staff health-related elements, as well as “job involvement” and “quality of care” (section S6 in Multimedia Appendix 1).

**Discussion**

**Overview**

This study presented the development, verification, and analysis of a CLD aiming to describe home care, including the impact of organizational change and reablement. The development process used participatory methods and qualitative verification to ensure fit-for-purpose.

**Model Analysis and Potential Impact**

The model showed great potential for facilitating discussions of knowledge in the home care domain. The activities supported its use for informing improvement, the study design of intervention studies, and future quantitative modeling. For example, by using a systems approach, key elements, their connections, and accompanying indicators and instruments, can be mapped before the design of a study. The model could also serve as a basis for discussing organizational improvement and how to best plan care and services to meet demand while minimizing staff workload, considering the full complexity of the system. This is similarly to how CLDs have been used in health research to inform improvement and policymaking for health promotion, mental health, health systems, and combating antimicrobial resistance [18,22,25,120].

Important leverage points of the model were identified. These are potential targets for the design of interventions. For example, the model highlighted the importance of home care user “needs met” on both home care user, “adverse events or progression of disease,” and home care staff “provision of care and services” and “workload.” Meaning that this element is important for determining both home care staff’s “distress” and home care user “satisfaction with care,” “quality of life,” and “likelihood of institutional care.” Similarly to the job demands-resources model [116], home care staff “social support” indirectly had a positive effect on the “ability to cope” through a positive effect on “recognition,” therefore counteracting “distress.” This highlights
the importance of taking a holistic systems perspective on interactions between elements when studying health care–related systems.

Focusing on the key elements of interest in this study, the behavior following activation of “person-centered care” (ie, the reablement approach) supported current evidence on the reablement approach and its influence on home care user needs met, functional ability and autonomy, and staff workload. The adoption of technology resulted in improved communication and coordination between services, but also an increased “complexity of work tasks” and “equipment problems.” Technology in the context of the reablement approach can lead to several benefits. However, care should be taken to ensure that this does not increase the workload of home care staff. Activation of “workload” and “distress” were important in determining the health of home care staff and the provision and quality of care.

**Limitations**

The final model is subject to several limitations and assumptions and should not be taken as ground truth. Here, the participating expert group consisted of academic experts in health care sciences with a focus on home care and the reablement approach. To ensure the relevance of the model, it is of value to engage with stakeholders more broadly, including home care users, their relatives, home care staff, management, and policymakers. The model was also shaped by the user requirements and context under which it was developed. Here the main emphasis was on home care staff workload, stress, the impact of the reablement approach on the delivery of care and services, and other organizational factors. As multiple perspectives were explored during model development, the literature searches were not systematic reviews and should not be considered exhaustive, hence this introduces a potential source of bias. Due to the large size of the collected data set statistically significant and only quantitative relationships were considered for the model. Therefore, excluding potentially relevant effects. The model is grounded in the universal care system as relevant to Sweden. We can expect a higher relevance of the financial burden of provision of care and services on home care users and caring relatives in systems where home care is paid for by the users [121,122]. Hence, this work should not be viewed as a generic model of home care, although we believe it to be valuable for informing modeling work in other contexts.

Data on home care were supplemented with additional evidence from nursing home and residential care settings. To ensure model validity, the CLD was reviewed by the experts from the perspective of which setting evidence originated. The number of elements and relationships supported by evidence outside of the home care setting alone were minimal and still viewed to be of relevance to home care. Similarly, model verification was supplemented with intervention studies in the nursing home setting. Studies of interventions on staff communication skills, emotion-oriented care, staff de-escalation skills for aggressive behavior, and training on behavioral psychological symptoms for dementia may still be viewed as relevant for the home care setting [27]. While staff team building and supervisor training may be of less relevance due to the operational and organizational differences between home care and nursing homes [28,29]. Hence, this is an important consideration when interpreting the results. Nevertheless, as discussed above the difference between settings may vary between health systems and the overall results from the verification exercise (reablement studies in home care, intervention studies in nursing homes, and expert-based review) suggest the model be representative with regard to its aims. Future verification and validation will consider data originating from the home care setting in Sweden, based on the studies being carried out in the Future Care research program.

The qualitative simulations provided insight into the potential effect of modulating elements in the CLD. However, it should be noted that with equal weighting of all connections, this did not account for nonlinear effects such as the impact of demand and control on distress. This work can be further extended with more quantitative analysis methods. Using methods such as Boolean modeling or page rank, which were used in our previous work on a qualitative systems model of mental health [112], may be of value in further exploring the system behavior.

**Future Work**

During development, reviewing the model during participatory activities served as an important medium for reflection and discussion on improvement and research in the domain of home care, work-related stress, and reablement. The model captured several aspects relating to the broader Future Care program. Going forward, the model may find use for study design as it encourages systems thinking when designing indicator sets for study protocols. This is further aided by the collated data where researchers can look up relevant instruments for measuring outcome variables across the diagram. The model may also serve as a basis for quantitative analysis of study data using structural equation modeling and perhaps even ordinary differential equation modeling in case of longitudinal data. Indeed, further work will focus on combining the qualitative system dynamics model with observed data on workload in home care to develop a quantitative predictive model.

**Conclusions**

In this work we developed, verified, and analyzed a causal loop model of workload, work-related stress, delivery of care and services, and reablement in the home care setting. The model showed consistency across the comparator data set and may therefore be of value for informing improvement and intervention studies within the context of home care. Further work will focus on the wider inclusion of stakeholders in participatory activities to reduce the risk of bias. Translation into a quantitative model will also be explored using observed data.
Acknowledgments

The authors would like to thank the Future Care research program group for their valuable expert feedback during the participatory activities of this study. Further, the authors would like to thank Susanne Assander (Karolinska Institute) and Arsineh Boodaighian Asl (KTH Royal Institute of Technology) for discussions on the modeling and simulation approach. This study was a part of the research program Future Care – for Older Adults in Home Care and Care Homes supported by FORTE, Sweden (grant 2016-07089). ASD, A-MB, SG, and SM received funding from Future Care through their respective institutions. For more information on Future Care, please visit the project website [23].

Authors’ Contributions

ASD, A-MB, SG, JR, and SM contributed to the study design, analysis and interpretation of data, drafting of the work, final approval of submission, and agree to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary Material - Methods & Results.
[PDF File (Adobe PDF File), 5719 KB - humanfactors_v10i1e42283_app1.pdf]

Multimedia Appendix 2
Supplementary Table S4.
[PDF File (Adobe PDF File), 1375 KB - humanfactors_v10i1e42283_app2.pdf]

Multimedia Appendix 3
Causal loop diagram: connections between elements and their type (Table S5) and elements, category, and social network analysis metrics (Table S6).
[PDF File (Adobe PDF File), 298 KB - humanfactors_v10i1e42283_app3.pdf]

Multimedia Appendix 4
Matlab simulation script.
[PDF File (Adobe PDF File), 24 KB - humanfactors_v10i1e42283_app4.pdf]

Multimedia Appendix 5
Causal loop diagram figure.
[PDF File (Adobe PDF File), 112 KB - humanfactors_v10i1e42283_app5.pdf]

References


Abbreviations

- CLD: causal loop diagram
- ICT: information and communication technology
- QPSnordic: the General Nordic Questionnaire for Psychological and Social Factors at Work
- SDCS: Strain in Dementia Care Scale

©Adam S Darwich, Anne-Marie Boström, Susanne Guidetti, Jayanth Raghothama, Sebastiaan Meijer. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 30.06.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

Joel Fossouo Tagné1,2,3, BBiomedSc, MPharm, GC-RESIM; Reginald Amin Yakob4, BBiomedSc, MPharm, MPS; Rachael Mcdonald3,5,6, BAAppSc, PGDip, GCHE, PhD; Nilmini Wickramasinghe1,6,7, BSc, Grad DipMgtSt, MBA, PhD

1School of Health Sciences and Biostatistics, Swinburne University of Technology, Melbourne, Australia
2Centre for Health Analytics, Murdoch Children’s Research Institute, Melbourne, Australia
3MedTechVic, Swinburne University of Technology, Melbourne, Australia
4Pharmaceutical Society of Australia, Sydney, Australia
5Department of Nursing and Allied Health, Occupational Therapy, Swinburne University of Technology, Melbourne, Australia
6Iverson Health Innovation Research Institute, Swinburne University of Technology, Melbourne, Australia
7Epworth Healthcare, Melbourne, Australia

Corresponding Author:
Joel Fossouo Tagné, BBiomedSc, MPharm, GC-RESIM
School of Health Sciences and Biostatistics
Swinburne University of Technology
John Street, Hawthorn
Melbourne, 3122
Australia
Phone: 61 0412478610
Email: jfossouo@gmail.com

Abstract

Background: Adverse drug reactions (ADRs) may cause serious injuries including death. Timely reporting of ADRs may play a significant role in patient safety; however, underreporting exists. Enhancing the electronic communication of ADR information to regulators and between health care providers has the potential to reduce recurrent ADRs and improve patient safety.

Objective: The main objectives were to explore the low rate of ADR reporting by community pharmacists (CPs) in Australia, evaluate the usability of an existing reporting system, and how this knowledge may influence the design of subsequent electronic ADR reporting systems.

Methods: The study was carried out in 2 stages. Stage 1 involved qualitative semistructured interviews to identify CPs’ perceived barriers and facilitators to ADR reporting. Data were analyzed by thematic analysis, and identified themes were subsequently aligned to the task-technology fit (TTF) framework. The second stage involved a usability evaluation of a commercial web-based ADR reporting system. A structured interview protocol that combined virtual observation, think-aloud moderating techniques, retrospective questioning of the overall user experience, and a System Usability Scale (SUS). The field notes from the interviews were subjected to thematic analysis.

Results: In total, 12 CPs were interviewed in stage 1, and 7 CPs participated in stage 2. The interview findings show that CPs are willing to report ADRs but face barriers from environmental, organizational, and IT infrastructures. Increasing ADR awareness, improving workplace practices, and implementing user-focused electronic reporting systems were seen as facilitators of ADR reporting. User testing of an existing system resulted in above average usability (SUS 68.57); however, functional and user interpretation issues were identified. Design elements such as a drop-down menu, free-text entry, checkbox, and prefilled data fields were perceived to be extremely useful for navigating the system and facilitating ADR reporting.

Conclusions: Existing reporting systems are not suited to report ADRs, or adapted to workflow, and are rarely used by CPs. Our study uncovered important contextual information for the design of future ADR reporting interventions. Based on our study, a multifaceted, theory-guided, user-centered, and best practice approach to design, implementation, and evaluation may be critical.
for the successful adoption of ADR reporting electronic interventions and patient safety. Future studies are needed to evaluate the effectiveness of theory-driven frameworks used in the design and implementation of ADR reporting systems.

**KEYWORDS**

pharmacovigilance; adverse drug reaction; pharmacist; user-centered design; activity theory

**Introduction**

**Background**

The use of medications may result in adverse drug reactions (ADRs) that may increase the risk of patient morbidity and mortality. The timely reporting of ADRs to regulators may contribute to patient safety by facilitating information gathering on drug safety data [1]. Worldwide ADR-related hospital admission ranges from 3.6% to 15.6% [2,3]. In Australia, ADR-related hospital admissions are estimated at 7.2% to 11% where 50% of ADR-related hospital admissions are preventable [4-6]. Furthermore, in Australia, medication-related problems account for approximately AUD $1.4 billion (US $937,440,000) per annum, that is, 15% of the total Australian Pharmaceutical Benefits Scheme [7]. The reporting of ADRs to regulators is important for postmarketing surveillance, quality improvements, and drug safety research, but they are vastly underreported [8]. The challenges to ADR reporting include lack of awareness by health professionals (HPs) or consumers, lack of time or financial incentives by HPs, legal implications, attitudes of the reporters (eg, reduced motivation, and lack of efficient/user-friendly reporting systems for clinicians) [9-11]. Poor documentation and reporting between care providers or across health care settings are a major roadblock to patient safety from known ADRs [12]. Critical information regarding ADRs or serious health conditions (eg, COVID-19) may remain elusive to HPs that prescribe and disperse drugs or regulators who govern or approve new drugs [12]. Furthermore, we previously identified that there is also substantial interinstitutional variability in the standards of ADR reporting among individual primary health care facilities by conducting a scoping review [13]. Therefore, re-exposure to harmful drugs can be potentially avoided by improving health care systems and medication supply practices [14].

Digital documentation and reporting of ADRs currently occur within web-based reporting systems hosted by regulatory organizations (eg, the Therapeutics Good Administration or Surveillance of Adverse Events Following Vaccination In the Community) [15,16].

Although these existing reporting systems are structured and standardized, they are perceived as cumbersome and time-consuming to navigate [1,16]. Therefore, ADR reporting within these websites is disconnected from the needs of HPs, and immediate patient care-related activities may supersede the data request of external agencies [12,17]. If ADR reporting was refocused to meet the patient safety information needs of the HPs who recognize, treat, and encounter new ADRs at the point of care, clinicians may be more willing to document and report these harmful events [12,18]. Digital health interventions such as e-prescribing, e-medical records, digital health records, and health mobile apps have been introduced in the last decade [19,20]. These interventions can promote efficiency across health processes, enhance patient satisfaction, and save costs [8]. Therefore, the uptake of electronic medical records provides opportunities for ADR reporting to be integrated into point-of-care systems [17]. Despite the promise that such technologies hold for integrated patient care and safety, their uptake among HPs has been slow, and this is likely due to assumptions that govern their design [12].

A 2020 systematic review of interventions to improve ADR reporting concluded that there was scope to include community pharmacists (CPs) to improve ADR reporting [9]. These findings were also consistent in other reviews [16,19]. The implementation of digital systems to support reporting by care providers and designing systems within the clinician workflow have been highly regarded [16,19,21]. To date, knowledge gaps exist regarding the practice and reporting of ADRs reporting by CPs in Australia [18]. To our knowledge, only 1 previous study explored the knowledge and perspective of CPs toward ADR reporting in Australia and found that 43% (n=101) of respondents agreed that a lack of time within their professional practice limited their reporting of ADRs and 65% (n=150) agreed that remuneration would encourage them to report ADRs [18]. Integration of autopopulation features within the dispensing software was identified as an efficient way to facilitate ADR reporting by CPs [18].

**ADR Surveillance Systems**

The safety surveillance of medications may be passive or active [22]. Active surveillance systems systematically monitor particular patient encounters to seek information about ADRs (eg, artificial intelligence), whereas passive surveillance systems provide opportunity for point-of-care providers to confidentially and voluntarily report ADRs [22]. In Australia, the GuildCare (GuildLink) is a passive surveillance system available in Australian community pharmacies [18,23]. The system was released in June 2014, allowing CPs to record and report ADRs directly to the Therapeutic Goods Administration (TGA) [8,18,24]. Soon after it was released (ie, June to September 2014), there was an increase in the rate of ADR reporting via the Guildlink portal [8]. The TGA received ADR reports nearly as high as that for the entire year of 2013, suggesting the system may have been well received by CPs [18]. However, despite the positive start, the numbers declined again in 2015, indicating there may be a need for constant reminders to maintain ADR reporting rates, and continuing system evaluation requirements [18]. A systematic review of adverse event reporting information systems found wide variation in the variety and type of data collected [25]. In addition, these reporting systems did not report pilot testing to ensure there was succinct, user-friendly, relevant, and correct interpretation of electronic fields by care providers.
prior to their implementation [25,26]. Implementing new systems without pilot testing and refining may fall short of their expected goals due to systems’ architecture constraints or design failures that could have been identified and resolved prior to their final build [27]. User-centered design (UCD) is a framework that places users (eg, CPs) at the center of the design process from the initial stages of planning, designing the system requirements, evaluating, and deployment of the final product [28]. It involves the influence of end users during the design processes and has been shown to contribute to the acceptance, adoption, and success of systems [29]. The core principles of UCD include: (1) understanding and specifying the context of use, (2) specifying the user including the organizational requirements, (3) producing design solutions, and (4) evaluating designs against requirements (Multimedia Appendix 1) [28,29].

Poor uptake of adverse event reporting systems by HP can occur when the system is designed without or with limited clinician input while prioritizing organizational data needs [12,30]. To optimize both the effectiveness and usefulness of these interventions, usability and acceptance are essential.

To date, there have been no reporting on factors affecting the implementation and adoption of pharmacovigilance (PV) systems in Australian community pharmacies and other primary care settings [16]. A key challenge to the successful utilization of any new system lies in strategies that drive uptake and adoption [28]. A recent 2022 systematic review concluded that future interventions should include a comprehensive multifaceted approach to improve the quantity and quality of ADR reporting [19]. A comprehensive multifaceted approach includes incorporating digital technologies with additional strategies that specifically address the key factors of a behavioral change framework [19]. The use of a behavioral change framework to investigate ADR reporting has previously been described [31,32]. As newer innovations emerge and digital technologies continue to transform health care management, several barriers still remain [33]. A recent Australian study reported the use of a theoretical domains framework together with a technological intervention as a strategy to facilitate ADR reporting by clinicians in hospitals [19,31]. However, the perspective of pharmacists working in community pharmacies is lacking [18,21]. The benefits of such digital systems are presumed to follow logic, and assumptions are that end users and clinical settings will adapt to the new technologies [12]. After identifying such assumptions and the potential detrimental impact on patient safety, our objectives were to understand why ADR reporting is low among CPs and examine barriers to reporting within their existing systems. This paper then provides insight into the application of activity theory (AT) from the fields of human behavior and information - communication as a framework to inform the evaluation or design of user-centered ADR reporting systems.

**Methods**

**Overview**

This was an exploratory study, with the underlying epistemology stemming from a social-constructivist paradigm, as the goal was to understand the knowledge constructed through CP’s practice lens. The study was predominantly qualitative, with some quantitative data that served as descriptive statistics. The study was carried out in 2 stages. Stage 1 was to gain a deeper understanding of the problem of low ADR reporting among CPs through understanding the “users” (CPs) and their social or environmental milieu. The individual results of stage 1 have previously been published [34]. Given the decline in electronic reporting through the Guildlink portal as discussed above, stage 2 was to evaluate the usability of a commercially available ADR reporting system (GuildCare system) to understand what attributes and features facilitate or prevent reporting (submitted for publication). Purposive sampling was used to select eligible participants working in community pharmacies listed on the Pharmacy Guild of Australia and Health direct website. Participants agreed to participate in a 25- to 60-minute recorded virtual interview. As we have submitted stage 2 for publication and previously published stage 1 of this work [34], in this paper, we have focused on the research results used to develop and propose our ADR reporting design and evaluation framework.

**Ethics Approval**

Before conducting the interviews, all participants provided informed written consent to participate in the study and were advised that the provided information may be deidentified and used for publication. Participants’ demographic data were collected by using a self-administered questionnaire, which was attached to the consent form. All procedures followed were in accordance with Australia’s National Statement on Ethical Conduct in Human research (2018). The study was approved by the Swinburne University of Technology Human Research Ethics Committee (Ref: 20214304-6249).

**Stage 1: Understanding the Problem (Low ADR Reporting), User (CPs), and the Context (Community Pharmacy)**

A qualitative study with individual interviews was conducted with CPs working across Victoria, Australia, between April 2022 and May 2022. A semistructured interview guide was used to identify CPs’ perceived barriers and facilitators to ADR reporting. Because this research also sought to explore strategies to implement innovative technologies to facilitate ADR reporting, the task-technology fit (TTF) model offered guidance when developing the semistructured interview questions and categorizing identified themes [8].

Task characteristics refer to the attributes of a task that can be executed using information communication technologies for the purpose of satisfying work practice needs (eg, dispensing a prescription). Tasks can vary in a number of dimensions including task nonroutineness, task interdependence, and time criticality. The users’ workflow and environment are also key considerations when assessing the “Fit” [35].

Technology characteristics refer to the technology tool used by individuals in carrying out their tasks. The aspects of technology tools may influence technology utilization and users’ perceptions. The TTF considers the importance of fitting the functionality and attributes of technology used, to the demands imposed by individual needs. Technology tools can either be hardware or software [36].
Data were analyzed by thematic analysis. Themes were constructed from the CPs’ reported barriers and facilitators. Thematic analysis began once interviews were complete using NVivo 12 software (QSR International). Initially, open codes were generated inductively from the participants’ descriptions of their experiences in reporting ADRs and the barriers or facilitators to reporting. The final analysis for this study focused on the key themes constructed from the interviews and was subsequently mapped into the TTF model. Data concordance was verified by coauthors NW and RM, who are both experienced in public and digital health research. Key themes were discussed with the research team that included clinicians with expertise in the quality use of medicine and drug safety. Interviews concluded when no additional themes could be identified and mapped to the TTF theoretical framework.

Stage 2: Usability Evaluation of an Online Reporting System (GuildCare)

A structured interview protocol (Figure 1) was developed that was designed to evaluate both usefulness and satisfaction; the interview protocol leveraged think-aloud moderating techniques (assessing usefulness), retrospective questioning about user satisfaction, and administration of the System Usability Scale (SUS; assessing satisfaction).

The SUS is a flexible questionnaire designed to assess any technology and is relatively quick and easy to complete [14,37]. It consists of 10 statements that are scored on a 5-point scale, 1=strongly disagree to 5=strongly agree, with the final scores (after transformation of the scores) ranging from 0 to 100 [38]. A higher score may indicate better usability. As a general rule, a system that has a score above 68 has acceptable usability; a lower score means that the system needs more scrutiny and continued improvement [14].

Usability testing relied on participants’ verbal communication and virtual observation through screen sharing [39]. During the interview, participants were directed to complete an ADR report scenario using a semistructured interview protocol. Thematic analysis began once interviews were completed using NVivo 12 software and was performed by 2 members of the team. The key themes were discussed among the research team that included a pharmacist and an engineer with experience in digital health. Interviews concluded when no additional themes relating to the research question could be found.

Figure 1. Summary of the system usability testing approach.

Results

Stage 1

In total, 12 CPs were interviewed. The themes identified spanned both task and technology aspects of the TTF sociotechnical framework. From the data, we identified the theme “lack of time” as a barrier to ADR reporting, which is consistent with previous studies in community pharmacy [18,40]; however, by using the TTF model, we were able to further analyze this theme by contextualizing what different CPs generally mean when they say, “lack of time to report.”

When CPs reported “lack of time,” this was either the requirement to stop performing regular duties, for example, clinical tasks and attend to the ADR reporting process or they were referring to the prolonged duration when “completing a reporting form,” for example, a digital regulatory reporting form. Within the first context, “lack of time” may be considered as a dependent variable, influenced by environmental factors, for example, the work environment or lack of support staff. In the second context, the CPs referred to the cumbersome web-based reporting forms. The identified barriers and suggested intervention strategies to ADR reporting is divided into 2 broad categories, corresponding to components of the TTF and is listed in Textbox 1.
Community pharmacists’ reported barriers and facilitators to adverse drug reaction reporting aligned to the task-technology fit.

- **Barriers corresponding to the task-technology fit framework**
  - Task: Lack of knowledge to adverse drug reaction reporting, time constraints, lack of financial incentives, lack of organizational support for adverse drug reaction reporting, and preference to refer consumers to physicians.
  - Technology: Low awareness to adverse drug reaction reporting systems, fragmented reporting systems, and inadequate organizational IT infrastructure.

- **Facilitators corresponding to the task-technology fit framework**
  - Task: Enhancing community pharmacists knowledge and awareness of adverse drug reactions, environmental restructuring and financial incentives for adverse drug reaction reporting, education, and empowering consumer reporting.
  - Technology: Workflow-integrated adverse drug reaction reporting technology systems, feedback provision to community pharmacists on the reported adverse drug reactions, and promoting consumer adverse drug reaction reporting.

Stage 2

In total, 7 CPs participated in the usability study. The system was perceived to have above average usability (SUS 68.57). Despite this, the use of a structured approach to usability testing identified themes that would have been overlooked by the results of the SUS alone. For example, when observing CPs navigate the system, all participants struggled to begin the task (ie, the ADR report) when they initially logged-in to the system. Despite the presence of 3 dots on the main user screen to begin the report, participants felt it was not clearly visible and lacked clarity (Figure 2). When adding the suspected medication participants were unsure about the frequency field (Figure 3), that is, 1, 2, or 3 times ongoing:

> not sure whether these options refers to the initial medication dose regimen or maybe the number of doses that had been taken or even the number of times the ADR was experienced. [CP4]

In addition to this, participants also struggled to complete and submit the ADR report and were confused as to why the form could not be submitted. In 5 of 7 interviewees, participants were verbally guided by the moderator to review and search for potential compulsory fields with missing data input that could prevent the form from being successfully submitted. This was not self-evident to the participants, adding more time to complete the report. Design elements such as a checkbox, drop-down menu, free-text entry, and prefilled/autopopulated data fields were perceived to be extremely useful for navigating the system and facilitating ADR reporting. Identified and reported themes have been divided into barriers and facilitators (Textbox 2).
The study objective was to understand why ADR reporting is low among CPs and examine the barriers or facilitators to ADR reporting within the existing system. The knowledge of barriers and facilitators to ADR reporting may inform the design of electronic ADR reporting systems that are fully integrated within the CPs workflow. Ideally, such systems will be used by clinicians (CPs) to facilitate ADR documentation at the point of care. Furthermore, they will allow for information sharing with regulators, among care providers and across health sectors to prevent unintentional re-exposures of patients to harmful drugs. This study sought to address a methodological gap in the way that ADR reporting systems have previously been conceptualized, designed, and implemented. Poor usability can arise from existing adverse event reporting systems that have been designed at a distance from users, with limited clinician input, prioritizing organizational data needs [12,30]. Previous research in Australia have reported that CPs are simply not using these systems, as the act of reporting is perceived as secondary to clinical care delivery, and systems are time-consuming, cumbersome to use, and are not integrated into current electronic information systems [18,19]. These findings have also been reported across multiple international jurisdictions [41]. The vast majority of ADRs remain underreported and are not reflected in the current health data that are used by regulators, including research organizations that examine drug safety [19,21,42]. Preventable ADRs may go unaddressed to the detriment and cost of health consumers, health care systems, and taxpayers [12].

**Theory-Driven Intervention Design**

Prior research on ADR underreporting have suggested initiatives to improve reporting predominantly focusing on the users and

---

**Textbox 2.** Observed and participant-reported barriers and facilitators.

**Barriers**

- Navigations (accessing and submitting the report; workflow, eg, required multiple steps)
- Irrelevant data fields
- Minimum required data
- Lack of system integration (web-based vs within dispensing system)
- Lack of interoperability (sharing with allied health)
- Length (number of data fields/questions)

**Facilitators**

- Drop-down menu
- Auto-filled sections
- Search options (eg, medications)
- Combination of checkboxes, drop-down menu, and free-text entry
- Direct submission of report to the Therapeutic Goods Administration
- Succinct list and relevant to setting

---

**Figure 3.** Main screen to add suspected medication displaying frequency times.
have rarely scrutinized the systems in place [16,19]. The results of these studies suggest shortcomings that include poor user knowledge, lack of awareness, clinical priorities, incentives, and workplace culture [17]. Prior studies to improve ADR reporting have not questioned the data-centric orientation of electronic reporting systems and have not explored systems shortcomings, or proposed ways to redesign reporting systems to facilitate reporting, complement clinical care while meeting the data needs of regulators [9,19]. Globally, studies have discussed interventions to improve ADR reporting among health professionals; however, the suggested interventions are of a general nature [9,16], without an evidence-based theoretical framework or adequate assessment of the end user needs [31,32].

An intervention applied in 1 setting may not be appropriate for another health setting, and there is therefore a need for an evidence-based method to guide the selection and implementation of relevant interventions [29,32]. The behavior of HPs in reporting ADRs can be influenced by different factors, including individual characteristics and those that involve the external environment [43]. Therefore, it is necessary to understand ADR reporting behavior of CPs using a well-defined theoretical approach [9,24]. In our study, we found the application of the TTF (stage 1) to be useful in our data analysis and understanding of the problem [34]. Where previous studies have reported “lack of time” as a barrier to ADR reporting [18], we were able to apply context, to understand what different CPs meant by the common phrase “Lack of time to report.” Nonetheless, the first context “lack of time” was a dependent variable, influenced by environmental constraints, for example, a lack of support staff, while in the second context, the CPs were referring to the cumbersome and time-consuming process of digital reporting web forms. The UCD approach begins with gaining a clear and thorough understanding of the users and task analysis, including the context of use, which is key to the implementation and adoption of the system [29]. Failure to understand the fundamental needs of end users when developing ADR reporting interventions may lead to reduced system usage and negatively impact patient safety. The TTF model has been applied in health care settings where businesses require technology solutions [44].

**Limitation of TTF in ADR Interventions Design**

The use of the TTF theoretical model to support our research inquiry may have limited the exploration of other important factors. For example, in stages 1 and 2, some CPs made comments such as “I think doctors are responsible for ADR reporting” or “we can report, but I'm sure anyone can report, including customers.” While in our study we determined the uncertainty around the responsibilities for ADR reporting as a lack of knowledge associated with the “task” (TTF), it may also suggest a lack of task ownership or task responsibility by the CPs, which could impact intervention design and successful implementation. In his theory on systems of professions, Abbott argued that individuals of professions generally define their jurisdictions, that is, the link between a profession and its work, by claiming exclusive rights over particular tasks [45]. However, in Australia, ADR reporting is a task that is not exclusive to CPs per se; instead, this task is conferred upon them and other HPs by the regulators (eg, TGA). CPs are not exclusively responsible for undertaking ADR reporting; it is a shared task between CPs, doctors, allied health professionals, health consumers, and the public [8]. Furthermore, it is a voluntary act to be performed for the purposes of advancing drug safety knowledge and patient safety. Therefore, when considering interventions or designing new systems, it would be beneficial to explore what happens to tasks like ADR reporting that are shared and not specifically claimed by a professional group. The ideal situation for drug safety monitoring would be that all HPs claim responsibility for the task and report. However, there is a blur in the boundaries of task allocation. This may result in the potential for ADR reporting to be ignored across HPs including CPs and could be a reason why the ADR reporting rate declined again after the initially reported increase in 2014 when the GuildCare ADR system was released [18].

**AT as a Conceptual Lens of Analysis in ADR Reporting**

Given the limitations found in the TTF, we undertook a further review of the various sociotechnical theories that could encompass the multifaceted and dynamic contexts involved in human decision and ADR reporting. The alternative theories explored included theories of planned behavior, technology acceptance model, and unified theory of acceptance and use of technology model. While these theoretical frameworks explore human behavior and use intention [46,47], they do not assess human interaction within the entire work system (eg, teams and organizations) as discussed in the limitations of using the TTF framework (ie, task ownership associated with ADR reporting). However, following further exploration, the constructs of AT were found to be fitting. AT is a descriptive approach that explains human practices in the social context (Multimedia Appendix 2) [8]. This theory considers the viewpoints and behaviors of users in a social context, originally based on the work of Vygotski and the study of cultural-historical psychology [48,49]. The AT framework uses “activity” as the fundamental unit to study human interaction [29]. The activity (what people do) is reflected through actions as people interact with their environment, thus providing a richer analysis of the user’s needs, context, and the direct or indirect environment [29]. The components of activity include subject, object, tools, rules, community, division of labor, and outcomes [49].

Reflecting on the activity model (Multimedia Appendix 2), researchers, designers, and developers of electronic ADR reporting systems may define the different constructs as follows: AT can be used to understand the interaction among the subjects (HPs or consumers/patients) and the objects (activities and processes involved in documenting and reporting ADRs). The tools in this study are the reporting systems (eg, GuildCare reporting systems) used to record and report ADRs to the regulators or share information with members of the community. The rules guiding these activities include the organizational, jurisdictional, or federal laws regarding ADR reporting [50]. The community that takes part in these activities may include pharmacists and other health professionals, patients, and regulators. Within these activities, work (PV) is divided among the community [29,48].
Linking AT Within UCD

Human-computer interaction is a complex interdisciplinary field, concerned with design, implementation, and evaluation [29]. As such, we hope to make a novel attempt to operationalize AT as a theoretical lens for a UCD framework to support improvements or the development of new electronic PV interventions in pharmacy. UCD (Multimedia Appendix 1) begins with a thorough understanding of the needs and requirements of the users (CPs). Analyzing the interaction among potential users is also very important, and based on the UCD approach, establishing the context in which users may use the system should be defined at the beginning.

Using AT, the user needs and requirements can be investigated to provide a structured and richer understanding of the subjects’ (users) needs as well as their related activities (eg, ADR reporting or clinical tasks). These activities can then be separated into subjects, tools (intervention), and objects (outcome). A usable system not only understands the needs of the user but also understands a user’s situation (ie, the context and environment) [51]. Therefore, AT can help to examine the user’s environment, including their social or cultural milieu. The organization requirements can also be explored using the constructs from AT with the UCD framework, which may also be useful in evaluating acceptance [29]. Furthermore, the design addresses the whole user experience, not solely focusing on the usability of the system but also ensuring a positive user experience [51].

The user experience may be evaluated through the use of questionnaires and interviews that probes end user experiences after using a system [52]. In stage 2 (usability testing), collectively, the CPs perceived the system to have above average usability (SUS 68.57). However, through our structured approach, combining virtual observation, think-aloud, and retrospective probing with the SUS, we observed functional and user interpretation issues impacting user experience that would have been easily overlooked if we had simply relied on the SUS results or interviews (Textbox 2); for example, difficulties in accessing and submitting the reporting forms or confusion over the intent of data fields. Considering HPs already face time constraints, factors impacting their time may be perceived as an additional documentation burden, causing reduced adoption and affecting patient safety. Therefore, it is important to note that users may have different perceptions, understanding, and expectations of a system, which may affect how they interact with the system and may not always be reflected in surveys or interviews [51]. Furthermore, human activity is directly influenced by social, cultural, and historical context, which adds further complexity [53]. Applying AT to UCD may help provide more emphasis on the user’s interaction and requirements. This may also help to bridge any gaps by adapting contextual information to the user’s situation and needs [53]. Based on our findings, we propose a framework for leveraging AT within UCD in ADR reporting (Multimedia Appendix 3). The establishment of this framework may support the requirement stage (user or organizations) of UCD. It may allow stakeholders to gain a comprehensive understanding of the context and the user needs prior to system design (Multimedia Appendix 3, steps 1 and 2). Furthermore, the proposed framework may also be used during system evaluation and iterations.

Future Strategies to Improve ADR Reporting

There is great potential to leverage recent developments in digital technologies to improve ADR reporting [8,19]. Digital technologies are widely available in the areas of automation, data mining, and signal detection of ADRs [8]. For example, in Australia, an active vaccine safety surveillance system integrated with national surveillance networks was successfully linked with a cloud-based pharmacy vaccination recording system to develop an automated active vaccine safety surveillance system for community pharmacies [54]. This was introduced in response to the COVID-19 pandemic and automatically reports immunizations directly to the Australian Immunisation Register [54]. Furthermore, increased advancements have been made in the area of artificial intelligence and machine learning in the detection of ADRs, with 1 study showing an 80% success rate in automated ADR detection in the hospital setting [55]. However, these ADR surveillance systems are different from passive surveillance. These involve manually reporting ADRs and are dependent on behavioral changes from the clinician, organizational or workplace structures, and operational/IT infrastructures [22]. Previous interventions such as education, reminders, feedback, and so forth, have only been temporarily effective in improving ADR reporting rates with the effect diminishing substantially within 6-12 months after implementation [18,56-58]. Furthermore, these interventions may need continuous maintenance to improve ADR reporting rates, which may be time-consuming and expensive [19]. Studies investigating ADR underreporting have primarily focused on knowledge and attitudes, advocating for interventions targeting provider behaviors [17].

However, in practice, the successful implementation and adoption of a new technology often hinge on how well these systems are integrated into organizational and clinical practice, and whether they meet the needs and expectations of the end users [12,29]. Applying theory-driven and best practice approaches (eg, our proposed AT and UCD framework) to systems design, implementation, and evaluation may bring more rigor, robustness, and accountability to new ADR surveillance interventions.

Limitations

The findings reflect the activities and opinions of CPs working within the settings where we were able to conduct the study. CPs’ responses may have been shaped by the organizational context for reporting ADRs within the jurisdiction. We used purposive sampling that could have resulted in selection bias. The sample size may be seen as a limitation; however, there were varied opinions from many who do not regularly report ADRs, suggesting the strength of socially desirable bias may not be too strong. The study focused its inquiry using the TTF theoretical model, which may have limited exploration of other important factors, as discussed earlier. We spoke with CPs who had experience in other care settings (eg, hospital pharmacies). The generalizability of our findings to other clinical areas may be limited as information infrastructures, work organization,
culture, or environmental conditions and tasks vary across facilities and jurisdictions.

Conclusions

A tremendous opportunity exists to leverage recent innovations in digital technologies to improve ADR reporting by CPs. To ensure successful uptake, we recommend that future reporting systems are provider focused and user-friendly. Furthermore, these systems should be integrated within the clinical workflow, enabling documentation and information sharing with regulators, allied health providers, and consumers. A comprehensive and multifaceted approach to systems design, implementation, and evaluation may improve adoption and ADR reporting. Importantly, these approaches must allow for meaningful engagement with clinician-users in the design, evaluation, and implementation phases and should include observational methods to identify differences between the actual and perceived use of ADR reporting systems. The framework outlined in this paper offers an example of how a socio-technical framework and a UCD approach may be integrated in an iterative fashion throughout the different stages of the intervention-design-cycle to meet this need, from analysis to deployment. In the future, it will be interesting to evaluate the success of such a framework and other theory-driven intervention strategies in terms of ADR reporting rates, patient safety, and health outcomes.

Acknowledgments

This study was supported by the Department Health and Biostatistics, School of Health Sciences, and Iverson Health Innovation Research Institute, Swinburne University of Technology, Melbourne, Australia.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Human-centered design processes.

Multimedia Appendix 2

Structure of human activity in activity theory.

Multimedia Appendix 3

Proposed framework for leveraging activity theory within the user-centered design of ADR reporting systems.

References


Abbreviations

ADR: adverse drug reaction
AT: activity theory
CP: community pharmacist
HP: health professional
PV: pharmacovigilance
TGA: Therapeutic Goods Administration
TTF: task-technology fit
UCD: user-centered design

Please cite as:
Fossouo Tagne J, Yakob RA, Mcdonald R, Wickramasinghe N
JMIR Hum Factors 2023;10:e43529
URL: https://humanfactors.jmir.org/2023/1/e43529
doi:10.2196/43529
PMID:36826985

©Joel Fossouo Tagne, Reginald Amin Yakob, Rachael Mcdonald, Nilmini Wickramasinghe. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 24.02.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

https://humanfactors.jmir.org/2023/1/e43529
Diversity in Stakeholder Groups in Generative Co-design for Digital Health: Assembly Procedure and Preliminary Assessment

Pieter Vandekerckhove¹, MSc; Job Timmermans², IR, PhD; Antoinette de Bont³, PhD; Marleen de Mul¹, PhD

¹Erasmus School of Health Policy & Management, Erasmus University Rotterdam, Rotterdam, Netherlands
²Department of Military Management Studies, Netherlands Defence Academy, Breda, Netherlands
³Tilburg School of Social & Behavioral Sciences, Tilburg University, Tilburg, Netherlands

Corresponding Author:
Pieter Vandekerckhove, MSc
Erasmus School of Health Policy & Management
Erasmus University Rotterdam
Burgemeester Oudlaan 50
Rotterdam, 3062PA
Netherlands
Phone: 31 10 408 8555
Email: vandekerckhove@eshpm.eur.nl

Abstract

Background: Diverse knowledge and ways of thinking are claimed to be important when involving stakeholders such as patients, care professionals, and care managers in a generative co-design (GCD) process. However, this claim is rather general and has not been operationalized; therefore, the influence of various stakeholders on the GCD process has not been empirically tested.

Objective: In this study, we aimed to take the first step in assessing stakeholder diversity by formulating a procedure to assemble a group of diverse stakeholders and test its influence in a GCD process.

Methods: To test the procedure and assess its influence on the GCD process, a case was selected involving a foundation that planned to develop a serious game to help people with cancer return to work. The procedure for assembling a stakeholder group involves snowball sampling and individual interviews, leading to the formation of 2 groups of stakeholders. Thirteen people were identified through snowball sampling, and they were briefly interviewed to assess their knowledge, inference experience, and communication skills. Two diverse stakeholder groups were formed, with one more potent than the other. The influence of both stakeholder groups on the GCD process was qualitatively assessed by comparing the knowledge output and related knowledge processing in 2 identical GCD workshops.

Results: Our hypothesis on diverse stakeholders was confirmed, although it also appeared that merely assessing the professional background of stakeholders was not sufficient to reach the full potential of the GCD process. The more potently diverse group had a stronger influence on knowledge output and knowledge processing, resulting in a more comprehensive problem definition and more precisely described solutions. In the less potently diverse group, none of the stakeholders had experience with abduction-2 inferencing, and this did not emerge in the GCD process, suggesting that at least one stakeholder should have previous abduction-2 experience.

Conclusions: A procedure to assemble a stakeholder group with specific criteria to assess the diversity of knowledge, ways of thinking, and communication can improve the potential of the GCD process and the resulting digital health.

(JMIR Hum Factors 2023;10:e38350) doi:10.2196/38350

KEYWORDS

collaborative design; design methodology; stakeholder involvement; participatory design; digital health

Introduction

Background

Stakeholders such as patients, care professionals, and care managers are considered to play an important role in designing and creating digital health [1-4]. A widely used form of co-design that can involve a group of people to develop a digital health product is generative co-design (GCD) [5,6]. GCD is characterized by a collective creative process whereby knowledge is shared by stakeholders to develop a product or
service, such as digital health [7-12]. In a GCD process, stakeholders are more actively involved in the creative design process than in a more classical design process [10].

A wide variety of people who do not necessarily have a design background, such as patients, care professionals, and health policy makers, can be GCD stakeholders in a digital health project. For instance, content experts such as patients (often referred to as “users”) may improve the uptake of the output, as their needs regarding user guidance, specific reminders, and personal tracking will likely be better addressed [13]. Health policy experts may also contribute to digital health development. For instance, it has been suggested that their involvement during the COVID-19 pandemic has led to improved alignment between payers and care professionals, which may have contributed to the rapid uptake of digital health [14,15].

There are both theoretical and practical issues when involving different stakeholders in GCD. From a theoretical standpoint, GCD scholars hypothesize that the more the diverse stakeholders involve in a group in terms of diverse knowledge and ways of thinking, the better the GCD process [10]. However, this claim is not clearly explicated, which may be due to the conceptual challenges present, such as the lack of consensus on the definition of “stakeholder” and “involvement” [16]. For instance, how one defines involvement depends on how one views stakeholder representation, the time involved in the project, and whether the scope focuses on the project or a wider cultural change [16-18]. In addition, GCD is part of a larger research field known as participatory design (PD) [10]. In PD, specific values are upheld, including democracy, equalized power relations, mutual learning, and situation-based actions [16,19]. However, these values are not currently applied explicitly in the GCD stakeholder selection procedure. For instance, adhering to a democratic principle could mean that not only a hospital manager but also current and future users should be involved in the development process of digital health. However, criteria have not been proposed to justify the selection of ideal participants.

From a practical point of view, assembling a diverse stakeholder group to design digital technology may require more deliberation in the health care field than in other sectors because the interests of the diverse stakeholders may not be aligned. This may lead to practical challenges for stakeholders in gaining trust and managing multiple stakeholders and time pressure when involving patients and physicians [20-25]. However, design practice manuals do not address how to overcome these additional challenges when using GCD to develop digital health [11,26,27].

When tackling these theoretical and practical issues and involving stakeholders in the GCD process to develop digital health, there is little scientific guidance to help select the best stakeholders. No study has evaluated the performance of different stakeholder groups when using GCD to develop digital health. A meta-review, albeit limited to the development of serious games, has highlighted the need for this research, as the effect of involving some users as stakeholders in PD studies is unclear [28].

**Objective**

To provide further scientific guidance on the involvement of stakeholders, we tested the hypothesis that stakeholders with more diverse knowledge and ways of thinking would improve the GCD process. To satisfy this aim, we operationalized the hypothesis through a procedure to assemble distinct stakeholder groups and assess their influence on the GCD process and output. As such, the research question is as follows: Do stakeholders with diverse knowledge and diverse ways of thinking improve the GCD process for digital health? The study’s goal is to conduct a preliminary assessment of diverse stakeholder groups assembled through a prescribed procedure in the early stages of a GCD process of a digital health project. This assessment will hopefully provide deeper insights that other researchers and practitioners can consider when deciding the most appropriate stakeholder to involve in their GCD project. With time, this could lead to a validated GCD stakeholder involvement procedure for digital health.

**Methods**

**Procedure to Assemble Diverse Stakeholder Groups**

The stakeholder group assembly procedure amounts to the operationalization of the Sanders and Stappers [10] hypothesis that stakeholders with more diverse knowledge and ways of thinking could improve the GCD process. To involve stakeholders who meet these requirements in a GCD process, a procedure containing 3 steps was followed: snowball sampling, interviews, and assemblage of stakeholders (Figure 1).

First, to recruit people, one needs to identify those who are committed to addressing the problem at hand. It can be useful to sample stakeholders through relevant organizations, associations, or events [25,29]. This should help ensure their commitment to solving problems, as these people have directly or indirectly been exposed to the problems and are logically more motivated to develop a solution.

Second, individual interviews can be conducted to qualitatively assess the diversity of knowledge and ways of thinking of the potential members. To operationalize the term “knowledge,” we define 3 types of knowledge (Textbox 1) based on the work of Batens [30-32]. One key form of knowledge that is also defined in GCD research is the deeper-lying tacit knowledge [10], which we measure here as contextual certainties. In addition, there are methodological instructions and relevant statements. Each of these 3 types of knowledge was assessed during an interview on a scale of 0 to 3 (Table 1). Stakeholders with extensive knowledge regarding the relevant statements and contextual certainties will be given the maximum score (3); stakeholders who are uncertain are given a score of 2 and those who seemed to have little knowledge, or did not provide relevant information in the interview, were awarded lower scores (1 and 0, respectively).

To operationalize the other component, “thinking,” we define 4 types of inferences, namely, induction, deduction, abduction-1, and abduction-2 (Textbox 1), as categorized initially by Peirce [33,37,38]. In particular, abduction-2 inferencing is expected to play an important role in the design process [33,38] and is
typically attributed to how designers think. Previous experience with these types of inferences can be assessed during an interview by counting the number of times an inference is used (Table 1). Abduction-1 can be scored as the number of methodological instructions formulated as concrete solutions (e.g., having an overview of one’s energy capacity after cancer treatment to continue work). Abduction-2 can be scored by looking at the use of generative heuristics as analogies or metaphors.

**Figure 1.** Stakeholder group assembly procedure.

**Textbox 1.** Working definitions of knowledge and inference types used for assessment.

<table>
<thead>
<tr>
<th>Knowledge types</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Contextual certainties</td>
</tr>
<tr>
<td>• Knowledge containing a deeper-lying perspective or philosophical principle</td>
</tr>
<tr>
<td>• Methodological instructions</td>
</tr>
<tr>
<td>• An approach to solve a problem or subproblem such as a procedure for operations, instruments, or tools</td>
</tr>
<tr>
<td>• Relevant statements</td>
</tr>
<tr>
<td>• Factual knowledge about the problem or the solution</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Inference types</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Induction</td>
</tr>
<tr>
<td>• A sequence of reasoning steps leading to a generalization, whereby several similar utterances are grouped under a new term or name, often in the form of a remark or conclusion following the utterances of others [33]</td>
</tr>
<tr>
<td>• Deduction</td>
</tr>
<tr>
<td>• A sequence of reasoning steps leading to a conclusion based on several previous utterances [33]</td>
</tr>
<tr>
<td>• Abduction-1</td>
</tr>
<tr>
<td>• A sequence of reasoning steps leading to the suggestion of a solution in the form of a methodological instruction</td>
</tr>
<tr>
<td>• Abduction-2</td>
</tr>
<tr>
<td>• A sequence of reasoning steps leading to the suggestion of a solution in the form of a methodological instruction whereby induction, deduction, abduction-1 and generative heuristics can be used, for example, a metaphor [34,35] or analogy [36]</td>
</tr>
</tbody>
</table>
In addition, communication skills can be assessed to determine whether potential stakeholders can effectively communicate their ideas to others in a group. For instance, we can assess whether a patient has the appropriate content expert background with various relevant statements that they feel confident to share during a GCD process with other stakeholders by asking the respondent for a self-evaluation.

Third, after conducting the interviews and scoring the responses, a diverse stakeholder group can be assembled based on 3 criteria. One can start by combining people from different professional backgrounds. Next, one can ensure that those stakeholders with the highest knowledge scores are included as they have more knowledge. In other words, if there are 2 stakeholders with the same professional background, the one with the highest score is included. Finally, the diversity of inferencing experience can be assessed. Here, one should ensure that a stakeholder group covers all inference types. Once one is satisfied that the stakeholder group covers all inference types, one can seek out the stakeholders with the greatest inference experience. For instance, if there are 2 stakeholders with abduction-2 experience, the one with the most experience (highest score) can be selected.

### Action Research Approach

To assess the stakeholder group assembly procedure, an action research approach [39] was used to guide the practitioners of a GCD project while adding the stakeholder group assembly procedure to simultaneously gain research insights.

### Hypothesis to Test

The aim was to test how a stakeholder group, assembled using the stakeholder group assembly procedure described in the aforementioned section, would influence the GCD process. We expected that this stakeholder group assembly procedure would produce a group with diverse knowledge and ways and that this would have a positive influence on the GCD process and output. We also expected that, in such a group, the “contextual certainties” knowledge type would be expressed more often by all stakeholders and the “abduction-2” inference type would be more often used specifically by the stakeholders with design expertise than in our less-experienced comparison group.

### Digital Health Project

A digital health development project in which multiple stakeholders could be involved in the GCD process was sought, and we could test the stakeholder assembly procedure to determine if it could make the GCD process more methodologically sound. Given the expertise of the first author (PV) with the problems faced by patients with cancer, a related project was identified and initiated by a Dutch cancer foundation called oPuce (The Foundation). The Foundation aims to create awareness of the stigmatization of cancer and supports initiatives to help people with cancer continue working during and after the illness and promote their return to paid work [40]. The Foundation had planned to start the development of a serious game to help people with cancer address their work-related needs. Although the actual development process had not yet started, The Foundation was interested in using a co-design process to develop the serious game. Because The Foundation had a large network of people who could potentially be involved as stakeholders in the design process to develop the serious game, we chose to add the stakeholder group assembly procedure as a first step in this process and help them with the first GCD activity.

### Ethics Approval

Ethics approval was granted by Erasmus Medical Centre’s Ethics Committee (MEC-2021-0231).

### Assembled Stakeholder Group

#### Overview

The stakeholder group assembly procedure described in the aforementioned section was followed in this study. The research data were solely managed by the first author (PV). The stakeholders received no financial compensation to participate in this study.

Here, we describe how snowball sampling, interviews, and group assembly were carried out. The first author initiated the snowball sampling [41] by approaching people at The Foundation via email and phone to identify stakeholders. At the end of this process, 13 potential stakeholders who had been...
involved in the initial conversations over the development of a serious game were identified (Table 2).

The 13 potential stakeholders were each assessed through 45-minute interviews, except for the network coordinator with COVID-19. Before the interviews, the participants were informed about the research and asked for informed consent. The web-based audio and video recorded interviews were carried out by PV and facilitated by creative exercises on Miro’s web-based collaborative whiteboard platform (Miro Corp; Multimedia Appendix 1). The creative exercises helped the interviewees gain a visual understanding of their ideas and become accustomed to the web-based creative software they would use during the GCD workshop.

Given that there were multiple stakeholders with similar backgrounds but scored differently in terms of knowledge and inference, the stakeholders could be divided into 2 groups (Tables 3 and 4). A more potent stakeholder group was formed of stakeholders with diverse backgrounds who scored highly on the knowledge and inference criteria. These stakeholders scored high in terms of providing more relevant statements and contextual certainties. This group had experience with all the inference types. A less potent stakeholder group was formed of the remaining stakeholders who still met the desired range of diverse backgrounds but scored less on the knowledge and inference criteria by showing less extensive knowledge and less inferencing experience during the interviews. Notably, none of the stakeholders in this group had experience with abduction-2 inferencing.

The stakeholders in both groups were unaware of this selection procedure, or why they were placed in which group, and the detailed aims of the study.

**Table 2.** Number of potential stakeholders identified through snowball sampling per professional background (N=13).

<table>
<thead>
<tr>
<th>Background</th>
<th>Stakeholder, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Game developer and designer</td>
<td>1 (8)</td>
</tr>
<tr>
<td>Employer (employing people with cancer)</td>
<td>3 (23)</td>
</tr>
<tr>
<td>Employer network</td>
<td>2 (15)</td>
</tr>
<tr>
<td>Employed cancer survivor</td>
<td>1 (8)</td>
</tr>
<tr>
<td>Occupational physician</td>
<td>1 (8)</td>
</tr>
<tr>
<td>Researcher</td>
<td>3 (23)</td>
</tr>
<tr>
<td>Network coordinator and patient with a previous history of cancer</td>
<td>1 (8)</td>
</tr>
<tr>
<td>IT manager</td>
<td>1 (8)</td>
</tr>
</tbody>
</table>

**Table 3.** Scores of stakeholders in the more potent diverse group.

<table>
<thead>
<tr>
<th>Background</th>
<th>Score(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Game developer and designer</td>
<td>11</td>
</tr>
<tr>
<td>Employer (employing people with cancer in company A) and facilitator</td>
<td>11</td>
</tr>
<tr>
<td>Employer (employing people with cancer in company B)</td>
<td>9</td>
</tr>
<tr>
<td>Employer network</td>
<td>9</td>
</tr>
<tr>
<td>Employed cancer survivor</td>
<td>9.5</td>
</tr>
<tr>
<td>Occupational physician</td>
<td>10</td>
</tr>
<tr>
<td>Researcher</td>
<td>11.5</td>
</tr>
</tbody>
</table>

\(^a\)Average score per stakeholder is 10 (SD 0.95).
The influences of coded and analyzed using ATLAS.ti (Mac Version 22.1.0; Scientific Software Development GmbH). The data from the interviews and workshops were iteratively hypotheses and aims of the study.

2 stakeholders were asked if they would take on the double role workshops as a co-designer. In addition, before the workshops, participants were given information about the aim of the identical parallel-running workshops, and a link was provided to familiarize themselves with the web-based Miro platform. GCD workshops are social activities in which stakeholders can share knowledge and work with creative exercises toward achieving the purpose of the design project [10,42,43]. Web-based workshops were considered the best option given the COVID-19 pandemic restrictions. The 30-minute web-based GCD workshops were audio and video recorded.

To provide a focus for the assessments, the GCD workshops were slightly artificially divided into 2 phases: the problem phase with the aim to understand the issues to formulate a problem definition and the solution phase to create ideas for a solution. The materials used in the 2 parallel-running GCD workshops were identical and organized specifically to focus on the interactions among stakeholders in both phases. Both groups received 5 identical instructions with a hexagon template delineating both the problem and solution phases, and sticky notes were provided (Multimedia Appendix 1).

In terms of roles, PV similarly facilitated both workshops and switched between them to ensure that the instructions were clear while consciously avoiding steering the content development process. Each stakeholder participated in the respective workshops as a co-designer. In addition, before the workshops, 2 stakeholders were asked if they would take on the double role of a participant and an assistant facilitator. All participants, including the assistant facilitators, were blinded to the hypotheses and aims of the study.

Data Collection

Data were collected during individual interviews as part of the stakeholder assessment procedure. In addition, data were collected in 2 identical parallel workshops that were part of a larger web-based event organized by The Foundation regarding the working of their organization. Before the workshops, all the stakeholders were given information about the aim of the identical parallel-running workshops, and a link was provided to familiarize themselves with the web-based Miro platform. GCD workshops are social activities in which stakeholders can share knowledge and work with creative exercises toward achieving the purpose of the design project [10,42,43]. Web-based workshops were considered the best option given the COVID-19 pandemic restrictions. The 30-minute web-based GCD workshops were audio and video recorded.

To provide a focus for the assessments, the GCD workshops were slightly artificially divided into 2 phases: the problem phase with the aim to understand the issues to formulate a problem definition and the solution phase to create ideas for a solution. The materials used in the 2 parallel-running GCD workshops were identical and organized specifically to focus on the interactions among stakeholders in both phases. Both groups received 5 identical instructions with a hexagon template delineating both the problem and solution phases, and sticky notes were provided (Multimedia Appendix 1).

In terms of roles, PV similarly facilitated both workshops and switched between them to ensure that the instructions were clear while consciously avoiding steering the content development process. Each stakeholder participated in the respective workshops as a co-designer. In addition, before the workshops, 2 stakeholders were asked if they would take on the double role of a participant and an assistant facilitator. All participants, including the assistant facilitators, were blinded to the hypotheses and aims of the study.

Qualitative Analysis

The data from the interviews and workshops were iteratively coded and analyzed using ATLAS.ti (Mac Version 22.1.0; Scientific Software Development GmbH). The influences of

Table 4. Scores of stakeholders in the less potent diverse group.

<table>
<thead>
<tr>
<th>Background</th>
<th>Score^a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Researcher 1</td>
<td>5</td>
</tr>
<tr>
<td>Researcher 2</td>
<td>3.5</td>
</tr>
<tr>
<td>IT manager</td>
<td>2.5</td>
</tr>
<tr>
<td>Employer network</td>
<td>3.5</td>
</tr>
<tr>
<td>Employer and facilitator</td>
<td>6</td>
</tr>
<tr>
<td>Network coordinator and cancer survi</td>
<td>10</td>
</tr>
<tr>
<td>Ecosystem expert</td>
<td>-^d</td>
</tr>
</tbody>
</table>

^a Average score per stakeholder is 5 (SD 2.47).

^b No formal interview was conducted; information was gathered through informal conversations.

^c No interview was conducted because this stakeholder only joined as an observer at the start of the generative co-design workshop.

^d Not available.
inferences were coded and analyzed in an empirical design setting. As such, utterances that bore similarities to the logical inference forms were coded according to the appropriate inference type (Textbox 1).

To further qualify the knowledge processing and knowledge output identified with the above-described deductive codes, 17 inductive codes (Multimedia Appendix 2) were used to identify stakeholder behaviors (eg, suggest a new idea or a reformulation; Table 5). These were used to understand why certain knowledge or inference types were used in each sequence.

Table 5. Examples of inductive code names and definitions to assess changes of knowledge within the workshops (see Multimedia Appendix 2 for complete list).

<table>
<thead>
<tr>
<th>Code name</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduce</td>
<td>Utterance whereby a new idea is proposed</td>
</tr>
<tr>
<td>Reformulate</td>
<td>Utterance whereby a previous idea is expressed using different words</td>
</tr>
<tr>
<td>Add</td>
<td>Utterance whereby aspects are added to a new idea</td>
</tr>
</tbody>
</table>

To assess the knowledge output in a sequence during the solution phase, 4 inductive codes were used to code knowledge changes through stakeholder interactions (Figure 2): concrete specific (eg, proposing to use a coach), concrete general (eg, proposing to use artificial intelligence), abstract specific (eg, a virtual angel—a specific object or artifact), and abstract general (eg, an empowering journey—a general image that may contain several specific solutions).

Figure 2. Inductive codes to code the knowledge changes: on x-axis from abstract to concrete and on y-axis from general to specific.

Results

Main Findings

Our hypothesis on diverse stakeholders was confirmed, as the more potent stakeholder group had a relatively larger influence on the GCD workshop process and output in the problem phase (see Greater Processing of Relevant Statements Increased Knowledge About the Problem) and solution phase (see Greater Use of Abduction-2 Inferencing Improves the Concreteness and Specificity of Solutions) than the less potent group (Table 6). Regarding the problem phase, in terms of influence on the process, the more potent stakeholders built on each other’s relevant statements, some of which had already been mentioned in the interviews before the workshop. Here, we noticed a dual movement. On the one hand, there was an expansive movement of diverse knowledge as the varied stakeholders shared their knowledge about the problem, and on the other hand, there was a narrowing integrative movement in which the content of ideas changed, and this changed the course of the discussion. In terms of output, the more potent group developed a more comprehensive problem definition.

Regarding the solution phase, in terms of influence on the process, the more potent group used more abduction-2 inferences, leading to a greater variety of methodological instructions (Table 6). In addition, the more potent diverse stakeholder groups, as in the problem phase, developed each other’s methodological instructions. This made the solutions more concrete and specific. Therefore, in terms of GCD output in the solution phase, the more potent stakeholders had a greater influence, as this group produced more precisely described solutions.

The other 2 subhypotheses were not supported. Only once, and only implicitly, contextual certainties were identified in the GCD workshop (Table 6). This was true only among the more potent stakeholder groups. As such, there seems to be no substantial difference between the 2 groups in terms of explicitly sharing more tacit deeper-lying knowledge. Furthermore, although we had expected abduction-2 type inferencing to be applied by stakeholders with a design background, it was not
used by the game developer who was the only participant with this background in the more potent diverse stakeholder group. Rather, abduction-2 inferences were made by the nondesigners in this group, which is contrary to our expectations.

<table>
<thead>
<tr>
<th>Code group and code</th>
<th>Frequency in more potent group</th>
<th>Frequency in less potent group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Problem phase</td>
<td>Solution phase</td>
</tr>
<tr>
<td><strong>Inference type</strong></td>
<td><strong>10</strong></td>
<td><strong>0</strong></td>
</tr>
<tr>
<td><strong>Induction</strong> a</td>
<td><strong>9</strong></td>
<td><strong>6</strong></td>
</tr>
<tr>
<td><strong>Deduction</strong></td>
<td><strong>0</strong></td>
<td><strong>2</strong></td>
</tr>
<tr>
<td><strong>Abduction-1</strong></td>
<td><strong>0</strong></td>
<td><strong>13</strong></td>
</tr>
<tr>
<td><strong>Abduction-2</strong></td>
<td><strong>14</strong></td>
<td><strong>4</strong></td>
</tr>
<tr>
<td>Knowledge type</td>
<td><strong>0</strong></td>
<td><strong>24</strong></td>
</tr>
<tr>
<td><strong>Relevant statements</strong></td>
<td><strong>1</strong></td>
<td><strong>0</strong></td>
</tr>
</tbody>
</table>

aKey differences have been highlighted in italics.

**The Greater Processing of Relevant Statements**

**Increased Knowledge About the Problem**

In terms of interactions about the problem, the stakeholders in the more potent group shared a greater diversity of relevant statements (14 vs 10), which were processed using more induction (10 vs 0) and deduction inferences (9 vs 4) than the less potent diverse stakeholder group did (Table 6). Furthermore, the stakeholders in the first group built on each other’s relevant statements, some of which had already been mentioned in the interviews before the workshop. These interactions were related to focusing on the discussion, asking questions, explaining ideas, introducing new ideas, and reformulating old ones, which occurred more frequently in the more potent group.

How stakeholders in the more potent stakeholder group developed each other’s knowledge about the problem is clearly demonstrated in the examples of the more potent group (Table 7). The employer expanded the discussion concerning the self-management of cancer survivors and added that one should consider the resilience of these people and avoid putting them into a victim role. Although he had already mentioned the need for a bespoke resilient solution in the individual interview, this was not in relation to considering the victim role of a patient or in relation to self-management. The employer and facilitator reformulated these points slightly and responded that this comment was related to developing the content of the serious game rather than its implementation. The game developer specified (relevant statement) that these aspects concern the content and didactics behind the content of the serious game. This probably follows from a more abstract principle that the game designers believe in, that “the content of a serious game always has a didactic aim behind it” (contextual certainty). The employed cancer survivor returned to what the employer had mentioned earlier and questioned whether there was a victim role at all. Finally, the employer and facilitator attempted to integrate the different points and reformulate this as a new question.

Thus, in the more potent group, the stakeholders such as employers and a patient shared their views on the problem by asking questions, reformulating points, and trying to draw connections. They shared their different ways of viewing self-management for people with cancer looking forward to returning to work. As a stakeholder, the technological background of the game developer enabled him to quickly point out how this could be accommodated in a serious game through the underlying didactics. This shows how each of the different stakeholders in the OCD process can rapidly interject useful information to define the problem based on the actual needs while conforming to what is technically needed and possible.

The interaction between stakeholders in the less potent group (Table 8) was more a group conversation without people building on each other’s knowledge (relevant statements). This led to less integration of the knowledge that was being shared. Even though they seemed to make a start to focus on the aspect of the problem as “the barriers preventing people with cancer to resume work,” they did not ask each other what that means or attempted to define the barriers. In the more potent stakeholder group, we observed more concentrated attention on the content of the problem, which led to more integration of knowledge about the problem, for example, the concepts of self-management, the victim role, and serious game development were rapidly connected to a problem definition.
Table 7. Sequence with codes from more potent diverse stakeholder group (translated into English for reporting purposes).

<table>
<thead>
<tr>
<th>Stakeholder and sequence of utterances (order of conversation)</th>
<th>Behavior code</th>
<th>Inference-type code</th>
<th>Knowledge-type code</th>
<th>Repetition code</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Employer</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. It feels to me that a user-centered bespoke solution is very general. I mean, doesn’t that apply to any situation?</td>
<td>Focus</td>
<td>Deduction</td>
<td><em>a</em></td>
<td></td>
</tr>
<tr>
<td><strong>Employer and facilitator</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. How would you make it more concrete?</td>
<td>Focus and ask</td>
<td>Deduction</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Employer</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. For example, coming back to what was said previously, how can we facilitate self-management? How can we avoid creating a victim role?</td>
<td>Introduce</td>
<td>—</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Because we want to make something bespoke. For example, how can you contribute to the resilience of the candidates looking for work or those who want to maintain work?</td>
<td>Explain</td>
<td>Deduction</td>
<td>Relevant statement</td>
<td>From interview</td>
</tr>
<tr>
<td>It’s in line with self-management, but a bit more.</td>
<td>Reformulate</td>
<td>Induction</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Employer and facilitator</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. How can you connect that to a serious game? It’s obviously also a general problem.</td>
<td>Ask</td>
<td>Deduction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>How do you maintain self-management? How do you prevent the victim role? Then, you are in the development process of the serious game.</td>
<td>Reformulate</td>
<td>Induction</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Game developer and designer</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. But more content, the didactics behind it.</td>
<td>Introduce</td>
<td>Induction</td>
<td>Relevant statement and contextual certainty</td>
<td>From interview</td>
</tr>
<tr>
<td><strong>Employer</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. The content</td>
<td>Reformulate</td>
<td>Induction</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Game developer and designer</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Yes, indeed</td>
<td>Agree</td>
<td>—</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Employed cancer survivor</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. If there would be a victim role?</td>
<td>Ask</td>
<td>—</td>
<td>Relevant statement</td>
<td></td>
</tr>
<tr>
<td><strong>Employer and facilitator</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. I am thinking about the last point of (employer) and from (researcher) to keep it concrete and small and still also connect it with the piece on implementation.</td>
<td>Focus</td>
<td>Induce</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Then we arrive again at the point of how do we make sure that the serious game offers added value for individual employees with cancer, but then we still remain with a big problem.</td>
<td>Reformulate</td>
<td>Deduce</td>
<td>—</td>
<td></td>
</tr>
</tbody>
</table>

aNot available.

Over time, the interactions about the problem in the GCD workshop with the more potent stakeholders showed a dual movement that was not present in the less potent group. On the one hand, there was an expansive movement of diverse knowledge as the stakeholders shared more knowledge about the problem and on the other hand, there was a narrowing integration movement whereby the content of ideas changed, which changed the course of the discussion. For example, initially, there was an expansive diverse knowledge movement as various stakeholders discussed the broad theme of user-centeredness. Then, there was a narrowing integration discussion about the definition of the user, whereby the question was raised as to whether one should focus on the development or implementation aspects. Some aspects were considered together, as it was mentioned that self-management was important for users. Here, the initial ideas changed as this was rephrased to clarify that some aspects are relevant during the development phase of the serious game and others during its implementation. Other elements that were discussed concerned resilience and the victim roles to be considered (Table 8), although these were not integrated into the problem definition. This dual movement may have contributed to the more potent diverse stakeholder group having a more comprehensive problem definition (Textbox 2) than the less potent group. In the problem definition phase, the less potent stakeholder group seemed to have brought together ideas in an expansive
movement; however, there was no subsequent integration of or change in the content that formed the problem definition. The more potent group’s more elaborate problem definition seems to have provided a better-founded basis on which to develop solutions.

### Table 8. Sequence with codes from less potent diverse stakeholder group (translated into English for reporting purposes).

<table>
<thead>
<tr>
<th>Repetition code</th>
<th>Knowledge-type code</th>
<th>Inference-type code</th>
<th>Behavior code</th>
<th>Stakeholder and sequence of utterances (order of conversation)</th>
</tr>
</thead>
</table>
| From interview  |                     |                     |               | **Researcher 1**
|                 |                     |                     |               | 1. If I am now looking. I am focusing on the serious game. That seems to be the starting point. Then, I think a central problem is that we see that the current ways of people getting back to work are not successful. And we want to improve that. Improve self-management. Well, let’s continue here, I am sure you can add to this. |
|                 |                     |                     |               | **Employer and facilitator**
|                 |                     |                     | Ask           | 2. Does everyone agree? |
|                 |                     |                     |               | **Network coordinator and cancer survivor**
|                 |                     |                     | Introduce     | 3. I think also, how can you improve the collaboration? How can you, with each other? Perhaps intercompany or inter-academic? Perhaps, this has nothing to do with… |
|                 |                     |                     |               | **Ecosystem expert**
|                 |                     |                     | Introduce, reformulate, and ask | 4. What I thought is that solution-oriented thinking is more on the outside of the hexagon (exercise template). I think that the word removing barriers to resume work, that is for example a problem related to the content. I don’t know how others are looking at this? |
|                 |                     |                     |               | **Researcher 2**
|                 |                     |                     | Agree         | 5. I agree with that. |
|                 |                     |                     |               | **Network coordinator and cancer survivor**
|                 |                     |                     | Ask and reformulate | 6. This is about keeping your work? |
|                 |                     |                     | Deduction     | **Ecosystem expert**
|                 |                     |                     | Agree         | 7. Keeping your work. |

**Textbox 2. Problem definitions.**

**Problem definition of the more potent diverse stakeholder group**
- How do we realize a bespoke approach and self-management during the implementation of the serious game (whilst taking this into account during development of the serious game)?

**Problem definition of the less potent diverse stakeholder group**
- Maintaining work during and after cancer

**Greater Use of Abduction-2 Inferencing Improves the Concreteness and Specificity of Solutions**

In the solution phase, the more potent group of diverse stakeholders used more abduction-2 inferences (13 vs 0), which led to a greater variety of methodological instructions (24 vs 8) than those observed in the less potent group (Table 6). In addition, similar to what the stakeholders did in the problem phase, the more potent diverse stakeholder group developed each other’s methodological instructions in the solution phase. This resulted in more concrete and specific solutions. Furthermore, abduction-2 inferencing was used by nondesigners, which was less anticipated because inferencing is typically attributed to designers.

How stakeholders developed ideas based on each other’s methodological instructions and how this made the solution more concrete and precise are clearly demonstrated in the example of the more potent group (Table 9). The researcher suggested a solution that he explained as being a tool for a social network, using a Star Trek metaphor by referring to The Borg. This is an abstract solution, characterized by a metaphor, yet sufficiently specific, as it is further described as a social network. Next, other suggestions, each using a different
metaphor, were used as analogies to highlight different features or aspects of the social network. Thus, the solution became more concrete and specific. The occupational physician suggested a buddy system; the researcher suggested a similar swipe function as in a Tinder app; and the employer and facilitator suggested offering personal suggestions based on an artificial intelligence algorithm. The metaphors that were used seemed to come from popular culture or daily use, which may have made them immediately clear to all stakeholders. As such, the solution-related knowledge of the various stakeholders started on an abstract-specific level and moved toward a more concrete and specific level (Figure 3). Overall, the more potent diverse stakeholder group had a strong influence on the quality of the knowledge output regarding the solution.

The interaction in the less potent group was more on the level of sharing relevant statements about a solution, for example, improving the skills of people with cancer (Table 10). They did not discuss in more detail how skills training could be implemented with, for instance, visual images (abduction-2). Therefore, the solutions did not change from abstract to concrete; instead, they remained relatively the same at a concrete level.

Table 9. Example sequence utterances from the more potent diverse stakeholder group in the generative co-design workshop with codes (translated into English for reporting purposes).

<table>
<thead>
<tr>
<th>Stakeholder and sequence of utterances (order of conversation)</th>
<th>Behavior code</th>
<th>Inference-type code</th>
<th>Knowledge-type code</th>
<th>Repetition code</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Researcher</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. You are not as an individual… because in such a game you are addressed as an individual, so how do we keep the social element and your environment? As an image I have The Borg, that’s from Star Trek, and you are being assimilated in a very large network of other individuals.</td>
<td>Introduce</td>
<td>Abduction-2</td>
<td>Methodological instruction</td>
<td>From interview</td>
</tr>
<tr>
<td><strong>Game developer and designer</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. I didn’t know you were a Trekkie.</td>
<td>Joke</td>
<td>_</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td><strong>Researcher</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Wait until you see my costume, ha-ha.</td>
<td>Laugh</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td><strong>Occupational physician</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. I am thinking about a sort of buddy system, rather than peers with similar experience, use buddy’s to play together.</td>
<td>Introduce</td>
<td>Abduction-2</td>
<td>Methodological instruction</td>
<td>—</td>
</tr>
<tr>
<td><strong>Researcher</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Yes, and maybe we can therefore also connect that with a Tinder app, because which buddy would you like?</td>
<td>Introduce</td>
<td>Abduction-2</td>
<td>Methodological instruction</td>
<td>—</td>
</tr>
<tr>
<td><strong>Occupational physician</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Ha-ha.</td>
<td>Laugh</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td><strong>Employer and facilitator</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. And, there, the artificial intelligence rises to the surface again? So that you can see on the basis of your use of the game with who you have the best connection?</td>
<td>Introduce</td>
<td>Deduction and abduction-2</td>
<td>Methodological instruction</td>
<td>—</td>
</tr>
<tr>
<td><strong>Occupational physician</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Exactly.</td>
<td>Agree</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td><strong>Employer and facilitator</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. That you are not only swiping, but also get a suggestion, like Hi, this person could fit with you.</td>
<td>Explain</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

aFirst visual image.
bNot available.
cSecond visual image.
dThird visual image.
eFourth visual image.
Table 10. Example sequence utterances from the less potent diverse stakeholder group in the generative co-design workshop with codes (translated into English for reporting purposes).

<table>
<thead>
<tr>
<th>Stakeholder and sequence of utterances (order of conversation)</th>
<th>Behavior code</th>
<th>Inference-type code</th>
<th>Knowledge-type code</th>
<th>Repetition code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Employer network</td>
<td></td>
<td></td>
<td></td>
<td>From interview</td>
</tr>
<tr>
<td>1. I am still thinking about an approach including skills, how that would enable people. I put it left under (in Miro), I lost it…</td>
<td>Introduce</td>
<td>Abduction-1</td>
<td>Methodological instruction and relevant statement</td>
<td></td>
</tr>
<tr>
<td>Network coordinator and cancer survivor</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. No, but skills are really important. Here, you have to do something completely different, and you are looking at work differently.</td>
<td>Agree and add</td>
<td><em>a</em></td>
<td>Relevant statement</td>
<td>—</td>
</tr>
<tr>
<td>Ecosystem expert</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. But I think that next to the work environment also, if you assume that that was the work environment where you were, the other one could then call a different work environment. Then those skills arise again, because you can perhaps get the possibilities to develop yourself differently.</td>
<td>Add</td>
<td>Deduction</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Employer network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Yes, and when one conquers cancer, for example you have certain perseverance, that you are resilient. And when you focus on that, your employer can you help you realise this.</td>
<td>Add</td>
<td>Conclude</td>
<td>Relevant statement</td>
<td>—</td>
</tr>
</tbody>
</table>

*aNot available.

Discussion

Principal Findings

This study aimed to answer the following research question: *Do stakeholders with diverse knowledge and diverse ways of thinking improve the GCD process for digital health?* As a first step in attempting to answer this research question, we assessed how a diverse stakeholder group, put together using the proposed stakeholder group assembly procedure, would influence the GCD process. We also established a second stakeholder group consisting of individuals who scored less well in the preliminary interviews held to assess the required competencies.

Our preliminary findings confirm Sanders and Stappers’ main hypothesis that a group of stakeholders with diverse knowledge and ways of thinking has a positive influence on GCD. The more potent of the 2 diverse stakeholder groups had a relatively larger influence on the GCD workshop process and output. The stakeholders in the more potent group built more on each other’s
knowledge, which led to a more comprehensive problem definition and more precisely described solutions. In the problem phase, the stakeholders in the more potent group shared a greater diversity of relevant statements (14 vs 10), which were processed using more induction (10 vs 0) and deduction inferences (9 vs 4) than the ones in the less potent diverse stakeholder group. Furthermore, the stakeholders in the first group built on each other’s relevant statements, some of which had already been mentioned in the interviews before the workshop. This resulted through a dual movement toward a more comprehensive problem definition. In the solution phase, the more potent group of diverse stakeholders used more abduction-2 inferences (13 vs 0), which led to a greater variety of methodological instructions (24 vs 8) than those observed in the less potent group. In addition, similar to what the stakeholders did in the problem phase, the more potent diverse stakeholder groups developed each other’s methodological instructions in the solution phase. This resulted in solutions that were developed from a more abstract and general level toward a more concrete and specific level.

The other 2 subhypotheses were not supported. First, there was no substantial difference between the 2 groups in terms of explicitly sharing deeper-lying knowledge (contextual certainties). One contextual certainty was used implicitly in the more potent group. Second, abduction-2 inferences were used 13 times by nondesigners in the more potent group but not by the game designer in the more potent group. This result was contrary to our expectations.

Using a person’s professional background as the sole criterion for group member selection as, for example, done by Trischler et al [44], may not deliver the full potential of a GCD session. Rather, it is the combination of stakeholders with diverse and complementary knowledge in terms of 3 knowledge types (relevant statements, methodological instructions, and contextual certainties) and the most diverse and complementary inference experience in terms of 4 complementary inference skills (deduction, induction, abduction-1, and abduction-2) that enhances the GCD process and its output. Moreover, abduction-2 inferencing did not occur spontaneously in our study in the less potent diverse stakeholder group. Therefore, the involvement of at least one stakeholder with abduction-2 experience (not limited to professional designers) could be critical when using GCD in hierarchical hospital settings [25], with stakeholders who are not naturally involved in creative activities.

Furthermore, the speed brought about by the dual movement of divergence and convergence [45] in the problem phase could be due to the diversity of knowledge and thinking among the stakeholders, as each one has the potential to convergence or diverge. Here, each has knowledge that others lack and cannot think in ways that others can. In the problem phase, the example provided was about an idea that was rapidly considered from a patient experience and from the employer and technical development perspectives. This led to reformulations and the raising of new questions, which steered the process in a new direction. This could be viewed as a change of frame, or perspective, brought about through the interaction of different stakeholders. Although there is extensive literature on the framing process [46-49], the interactions of diverse stakeholders in the framing process have not yet been explicitly described. The example we provided in the solution phase suggests that framing involving diverse stakeholders can be viewed as a knowledge process that looks for a solution from different knowledge contexts that provide different perspectives when looking at a possible solution. During this process, we observed an implicit negotiation process, which has been mentioned by other researchers [47,50], in the sense that the stakeholders’ responses to the proposed solutions varied. On some occasions, stakeholders laughed, which may signify acceptance of a solution. This was surprising and unexpected given that it did not relate to their own knowledge context. As such, a stakeholder group with diverse knowledge and ways of thinking may be the most effective when it can reframe ideas rapidly.

The framing process may be accelerated when stakeholders share more contextual certainties. However, we observed only 1 event in the problem phase that demonstrated how a contextual certainty can rapidly bring a new perspective to a discussion; in this case, a didactic perspective that is essential when developing serious games [51,52]. This emphasizes the need to share deeper-lying knowledge in the GCD process [10] and the need to elucidate how they are used by different stakeholders in design theory more broadly [53]. The limited expression of contextual certainties in our study may be due to the lack of priming exercises [8] ahead of our workshops, coupled with the time pressure and workload of participants. This may have suppressed the participants’ awareness of deeper-lying ideas. This suggests that there may be a minimum critical time before people can share such deeper-lying knowledge that our workshops failed to exceed.

**Implications**

Finally, we reflect on our stakeholder group assembly procedure in light of the normative values present in the GCD that originate in the PD field [10]. In PD, broadly defined values are upheld such as democracy, equalized power relations, mutual learning, and situation-based actions [16,19]. Given the lack of theoretical consensus, there are no solid normative grounds on which to judge our stakeholder selection procedure. For instance, the democratic principle might imply that one should involve people who are affected by the design decisions made or the end product [19]. In addition, it is emphasized that power relations should be equalized by giving voice to those who may be invisible or weaker [16]. In terms of digital health, this could imply that patients and informal caregivers should be involved. As it is often difficult to get involved in a health care setting [21], we considered the use of a snowball sampling method. This is potentially more inclusive and faster than a widely advertised recruitment strategy that may not attract susceptible groups. As such, in the protocol, we tried to cast a wide net of possible participants through snowball sampling to include people and other vulnerable populations. However, to participate in and contribute to the GCD process, individuals should be able to bring new or complementary knowledge and inferencing experience to the stakeholder group. On the basis that they lacked these assets, we did not include cancer survivors in the more potent diverse stakeholder group, even though they were in a susceptible position. Furthermore, it is argued that
democracy requires educated and engaged people acting in their own interests and in the interest of the common good [54,55]. Kensing and Greenbaum [55] state that, when necessary, this should involve educating people in terms of the required technical jargon and engaging them in the process, an aspect related to the principle of mutual learning [16,19,55]. In this respect, Kleinsmann argues that in collaborative activities, there should be minimal shared understanding [56]. In our protocol, we tried to ensure this by looking for people with a basic interest in the topic through snowball sampling and then using self-assessment to evaluate group communication abilities. In this sense, we believe that the stakeholder group assembly procedure that we used can serve as an example of how these values can be respected while improving the GCD process and output.

Limitations
The designed stakeholder group assembly procedure was operationalized in a minimally viable form to meet the aim and scope of this study. Although the assessment process was intended to accurately score the knowledge, inference skills, and communication skills of potential group members, there may be a built-in bias in the questions. Although we attempted to limit this by discussing the formation of the groups within the research team, there may still be some errors in allocating individuals to one of the 2 groups.

Indeed, not all the criteria were sufficiently sensitive to differentiate between the experiences of some stakeholders to ensure robust selection. For instance, all the stakeholders scored similarly on the criteria addressing induction and deduction inference types and communication abilities. This could be due to the snowball sampling that preselected stakeholders who were already part of The Foundation’s network with a certain level of educational training and communication abilities. Even though all the stakeholders showed a similar ability to use induction and deduction inference types in their interviews, the stakeholders in the less potent group used these less often during their workshop, which affected their knowledge output and knowledge processing. It is possible that the stakeholders in this group were less inclined to use these inference types because of a lack of interaction.

The case was selected based on the background of the lead researcher and the fact that it was a project that had momentum, was about to start, and had good potential to involve various stakeholders. However, the selected case also raised concerns, as it took longer than expected to gain approval to start the stakeholder selection procedure from the project manager. One reason for this could be that GCD is often used as an informal design practice rather than as a formal scientific approach with formal stakeholder selection.

We would caution readers against drawing any causal relationships based on our study about the influence of the stakeholder groups on the GCD process. To maintain focus in our analysis, back-and-forth interactions between the problem and solution phases, which might occur when addressing a real issue, were not considered. Furthermore, given the exploratory purpose of this study, various variables were ignored, including content-related facilitation, interpersonal relationships [57], the creative environment [58], mutual learning over time, and the higher-level strategy of the project and host organization [56,59]. Nevertheless, even without these aspects, this study was still able to provide initial insights into the role of stakeholder diversity in GCD. To ensure this, reflection meetings were organized between the lead researcher and coauthors to identify and avoid any potential biases in the study design and interpretation of the results.

Further Research
We would recommend further exploring how to strike a balance between the time and resources spent on snowball sampling and the number of stakeholder assessment criteria (knowledge, inference experience, and communication abilities) used. One option would be to ignore induction and deduction and focus on abduction-1 and abduction-2 inference experiences. One could also ignore communication abilities if the organization under consideration is a hospital that already requires interdisciplinary collaboration and focus instead on visual communication skills and open-mindedness as an indication of creative thinking. Next, to further assess the influence of the selected stakeholders on the knowledge processing component, the role of metaphors (in abduction-2 inferencing) and contextual certainties could be explored. For instance, one could link the dual-processing theory of reasoning, which involves deeper unconscious knowledge processing based on intuition and experience, and the more conscious deliberated processing with different knowledge and inference types [60]. Finally, the knowledge processing and knowledge output could, over time, be further assessed in the GCD process, in which the expression of contextual certainties is considered alongside stakeholders’ learning processes.

Conclusions
A procedure to assess the diversity of knowledge, diversity of ways of thinking, and communication skills in assembling a stakeholder group that meets specific criteria may improve the potential of the GCD process and the resulting digital health. We would encourage the validation of our preliminary findings. Ultimately, this will help researchers make methodologically more robust decisions about stakeholder involvement and report them in an appropriate way, which will improve the scientific rigor of GCD science for digital health.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Interview and workshop template used on online Miro platform.
References


40. oPuc foundation. oPuc. URL: https://opuce.opuce-foundation [accessed 2022-08-12]


Abbreviations

GCD: generative co-design
PD: participatory design

©Pieter Vandekerckhove, Job Timmermans, Antoinette de Bont, Marleen de Mul. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 14.02.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Involving Health Care Professionals in the Development of Electronic Health Records: Scoping Review

Theresa Sophie Busse¹, Dr Rer Med; Chantal Jux², MSc; Johannes Laser³, BA; Peter Rasche¹,⁴, Dr Ing, Dr Rer Med; Horst Christian Vollmar¹, Dr Med; Jan P Ehlers³, Dr Med Vet; Sven Kernebeck³, Dr Rer Med

¹Institute of General Practice and Family Medicine (AM RUB), Medical Faculty, Ruhr University Bochum, Bochum, Germany
²School of Nursing, Saint Elisabeth Group GmbH, Catholic Hospitals Rhine-Ruhr, Herne, Germany
³Chair of Didactics and Educational Research in Healthcare, Faculty of Health, Witten/Herdecke University, Witten, Germany
⁴Faculty of Health Care, Niederrhein University of Applied Sciences, Krefeld, Germany

Corresponding Author:
Theresa Sophie Busse, Dr Rer Med
Institute of General Practice and Family Medicine (AM RUB)
Medical Faculty
Ruhr University Bochum
Universitätstraße 150
MA 1/158
Bochum, 44801
Germany
Phone: 49 23432 ext 27782
Email: theresa.busse@ruhr-uni-bochum.de

Abstract

Background: Electronic health records (EHRs) are a promising approach to document and map (complex) health information gathered in health care worldwide. However, possible unintended consequences during use, which can occur owing to low usability or the lack of adaption to existing workflows (eg, high cognitive load), may pose a challenge. To prevent this, the involvement of users in the development of EHRs is crucial and growing. Overall, involvement is designed to be very multifaceted, for example, in terms of the timing, frequency, or even methods used to capture user preferences.

Objective: Setting, users and their needs, and the context and practice of health care must be considered in the design and subsequent implementation of EHRs. Many different approaches to user involvement exist, each requiring a variety of methodological choices. The aim of the study was to provide an overview of the existing forms of user involvement and the circumstances they need and to provide support for the planning of new involvement processes.

Methods: We conducted a scoping review to provide a database for future projects on which design of inclusion is worthwhile and to show the diversity of reporting. Using a very broad search string, we searched the PubMed, CINAHL, and Scopus databases. In addition, we searched Google Scholar. Hits were screened according to scoping review methodology and then examined, focusing on methods and materials, participants, frequency and design of the development, and competencies of the researchers involved.

Results: In total, 70 articles were included in the final analysis. There was a wide range of methods of involvement. Physicians and nurses were the most frequently included groups and, in most cases, were involved only once in the process. The approach of involvement (eg, co-design) was not specified in most of the studies (44/70, 63%). Further qualitative deficiencies in the reporting were evident in the presentation of the competences of members of the research and development teams. Think-aloud sessions, interviews, and prototypes were frequently used.

Conclusions: This review provides insights into the diversity of health care professionals’ involvement in the development of EHRs. It provides an overview of the different approaches in various fields of health care. However, it also shows the necessity of considering quality standards in the development of EHRs together with future users and the need for reporting this in future studies.

(JMIR Hum Factors 2023;10:e45598) doi:10.2196/45598
**KEYWORDS**
user-centered design; electronic health records; electronic medical records; digital technology; technology development; stakeholder participation

**Introduction**

**Background**

The use of electronic health records (EHRs) is increasing worldwide [1,2]. It has been associated with improvements in health care quality and patient safety [3]. In international literature, different terms are used interchangeably to refer to electronic clinical documentation, such as electronic medical records, electronic patient records, or EHRs [4]. In this paper, we use the term EHRs to refer to different types of electronic documentation of patient health data. EHRs are digitized medical records used in clinical health care within an organization [5]. EHRs are linked to organizations (e.g., an EHR that is used by the staff in an intensive care unit [ICU] of a hospital), as opposed to personal health records. Personal health records are characterized by the fact that patients can manage them themselves and provide access to others [6]. EHRs can electronically gather and record both administrative and health-related information as well as store, transmit, and display information from various sources [7]. Traditionally, health-related information in EHRs includes a medical history and medication orders, vital signs, or laboratory results [3,8].

Administrative information may include age, sex, or International Classification of Diseases codes [9]. Depending on the context, EHRs include different submodules, such as medication display, the display of vital signs, or diagnostic information [10]. For example, different content is more critical for work in an ICU than for work in a palliative care unit. Depending on the context, there are EHRs specific to each area of medical care to ensure optimal documentational support [11]. It is useful to ensure that information can be transferred within the units of a hospital and between health care institutions. However, this diversity of records still poses challenges for interoperability [12].

In recent years, technological progress has led to extreme improvements in the field of EHRs in terms of design and functions [13]. EHRs can be used to minimize costs and workload with the help of shared, location independent, and clear documentation [14] and to improve collaboration and coordination between different professions and individuals [15]. In addition to the digitization of previously paper-based documentation, electronic decision support systems and the use of predefined clinical guidelines and standards can support quality improvement based on the latest health care knowledge [3,16].

However, the solitary implementation and use of EHRs in isolation will not guarantee that the quality of care improves. The literature suggests that EHRs with poor usability or functionality may have unintended consequences for their users and patients [17]. For example, the lack of adaption to workflows [18,19] and user needs [20], poor usability [21], and unstructured data sets in EHRs lead to high cognitive demands on users [22]. These aspects are associated with work-related stress, fatigue, and burnout for the main user groups of EHRs: nurses [23,24] and physicians [25-27]. Furthermore, poor usability has been associated with patient harm [28]. For example, it can make it difficult for health care providers to access necessary medical data for the treatment of patients or lead to misinterpretation of available data. This can lead to misdiagnosis, incorrect treatment, or unsuitable medication for a patient’s condition. This can put the patient’s health and well-being at risk [28]. Therefore, user acceptance is essential for successful implementation, actual use, and user satisfaction of EHRs [29]. Expected usefulness, technical concerns, technical problems, and expected workflow challenges can facilitate or hinder technology acceptance [30].

To promote international joint development projects, globally valid standards have been drawn up. For example, there is International Organization for Standardization (ISO) 9241-210, which focuses on the ergonomics of human-system interaction. It is an important standard for classifying and demanding usability engineering measures in product development processes such as the development of EHRs. Unfortunately, these standards are often only partially complied with, leading to the various abovementioned problems.

In addition, the involvement of users is necessary to adapt EHRs to the needs of health care professionals and to ensure their acceptance [31]. This is increasingly being addressed, resulting in an expansion of projects involving future users in EHR development. Existing reviews have focused on the involvement of users in technology development. In recent years, several reviews have been published to address the involvement of users in the development of different health-related technologies [32,33]. The focus of these reviews has been, on the one hand, on the involvement of different user groups, such as older people [34-39], people living with dementia [40], or patients with chronic diseases [41]. Specific to these groups is the fact that their cognitive and physiological characteristics must be addressed in the development of digital technologies. On the other hand, reviews cover different use cases of technologies such as mobile health [42], serious digital games for health promotion [43], or for the treatment of depression [41]. In addition, other reviews cover more general aspects of user involvement for the development of health-related technologies [32,33]. Despite the empirical evidence supporting the need for user involvement in the development and implementation of EHRs, this topic is largely excluded from the reviews. For example, in 3 reviews covering generic aspects of user involvement, no study focused on EHRs [32,33,44]. Different approaches such as participatory design or co-design are common practices to involve users in the development of new technologies. For example, on the one hand, participatory design actively and creatively involves both users and designers and thus includes different individual qualifications [31,45]. This approach can be defined as “...a process of investigating, understanding, reflecting upon, establishing, developing, and supporting mutual learning between multiple participants in...”
collective ‘reflection-in-action’. The participants typically undertake the two principle roles of users and designers where the designers strive to learn the realities of the users’ situation while the users strive to articulate their desired aims and learn appropriate technological means to obtain them” [45]. On the other hand, an approach such as co-design is defined as an “active collaboration between stakeholders in the design of solutions to a pre-specified problem” [46]. Although these 2 approaches are often used interchangeably and synonymously, they differ in how much choice is given to users in the development of a technology. It can be assumed that the participatory design approach gives users more influence than the co-design approach.

Aim

In existing studies, the design of the methodology varies depending on resources, time period, and technology. When using participatory design or co-design methods, methodological choices must be made [46]. For the planning of similar research projects and a sensible use of diverse methods, it is crucial to provide an overview. Within the framework of a scoping review, we therefore investigated which forms of user involvement have been used to date, under what circumstances, and with what results. The result can also be used to facilitate guidelines for the involvement of health professionals in the development of EHRs. The overview in the metadata table in Multimedia Appendix 1 [47-115] is intended to be particularly helpful in this regard. In addition to the range of possibilities, specific classifications can also be made as to which method is helpful and for which objective.

The review was guided by the question: “How are health care professionals involved in the development of EHRs?”

Methods

Overview

The Methods section is reported as recommended by the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 statement [33]. The presentation of this scoping review is based on the methodological specifications by Peters et al [116].

This methodology was developed using an a priori scoping review protocol [117]. The decision to use a scoping review methodology was based on the identification of a gap in current knowledge [117] and the need for an overview of different methods without any assessment. The result should be a narrative account with a focus on the different ways and methods of involving end users in the development of EHRs.

The methodology of scoping reviews is gaining popularity, particularly in the field of health care [118]. Whereas systematic reviews aim to synthesize collate empirical evidence on a focused research question and present the evidence from the reviewed studies [119], scoping reviews map the existing literature on a topic area [120]. In addition, scoping reviews provide a descriptive overview [121] and are therefore an appropriate method for addressing the research question.

This review aimed to provide an overview of the existing ways and methods of user involvement in the development of EHRs in the literature. The four specific objectives of this review were (1) to conduct a systematic search of the published literature for studies focusing on user involvement in the development of EHRs, (2) to present the characteristics and range of methods used in the identified manuscripts, (3) to explore the reported challenges and limitations of the methods, and (4) to make recommendations for the further development of the approach to the development of EHRs and to improve the consistency with which these types of studies are conducted and reported.

Planning for the review began in January 2021. The review was conducted and evaluated from June 2021 to April 2022. Four people were involved in carrying out the review (JL, CJ, SK, and TSB). JL and CJ had experience in conducting (scoping) reviews. CJ, SK, and TSB worked with prospective users to develop an outpatient EHR, an inpatient EHR, and a cross-sectoral EHR for pediatric palliative care with future users.

Eligibility Criteria

Studies were eligible for inclusion if they described the involvement of health care professionals in the development of EHRs. This explicitly included studies that examined a specific EHR. However, excluded studies focused on the general workload resulting from the use of different EHRs in different institutions or other parameters related to different EHRs, regardless of their design. Articles published in languages other than English were excluded. Manuscripts that described a process without performing it were excluded from the scoping review. Gray literature was not included because of the focus on research projects, although this was included in the scoping review methodology [122].

SK and JL formulated the inclusion and exclusion criteria and discussed them with TSB and CJ. The inclusion and exclusion criteria are listed in Textboxes 1 and 2.
### Textbox 1. Inclusion criteria.

<table>
<thead>
<tr>
<th><strong>Languages</strong></th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Publication period</strong></td>
<td>2011-2021</td>
</tr>
<tr>
<td><strong>Format</strong></td>
<td>Full text available</td>
</tr>
<tr>
<td><strong>Study design</strong></td>
<td>Empirical studies on the development of an electronic health record (EHR) or modules or submodules where health care professionals are involved</td>
</tr>
<tr>
<td></td>
<td>Needs assessment</td>
</tr>
<tr>
<td></td>
<td>Requirements testing or evaluation</td>
</tr>
<tr>
<td></td>
<td>Qualitative, quantitative, and mixed methods studies</td>
</tr>
<tr>
<td><strong>Forms of publication</strong></td>
<td>Papers published in a scientific journal</td>
</tr>
<tr>
<td><strong>Product</strong></td>
<td>EHR</td>
</tr>
<tr>
<td></td>
<td>Submodules or modules integrated into an EHR</td>
</tr>
<tr>
<td></td>
<td>Same EHR in different stages of development</td>
</tr>
<tr>
<td><strong>Development phases</strong></td>
<td>Pending testing or a major effectiveness study</td>
</tr>
<tr>
<td></td>
<td>Implementation</td>
</tr>
<tr>
<td></td>
<td>Evaluation</td>
</tr>
<tr>
<td><strong>Setting</strong></td>
<td>All settings in health and social care</td>
</tr>
<tr>
<td><strong>Participants in development process</strong></td>
<td>Health care professionals, even if other groups of people are involved</td>
</tr>
</tbody>
</table>
Textbox 2. Exclusion criteria.

<table>
<thead>
<tr>
<th>Languages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Languages other than English</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Publication period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before 2011</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract only or full text not available</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Study design</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reviews</td>
</tr>
<tr>
<td>Randomized controlled trials</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Forms of publication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study protocols</td>
</tr>
<tr>
<td>Conference papers</td>
</tr>
<tr>
<td>Gray literature</td>
</tr>
<tr>
<td>Books</td>
</tr>
<tr>
<td>Bachelor thesis, master thesis, or similar works</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision support systems</td>
</tr>
<tr>
<td>Personal health records</td>
</tr>
<tr>
<td>Other technologies (integrated apps)</td>
</tr>
<tr>
<td>Comparison of different electronic health records in one survey</td>
</tr>
<tr>
<td>Electronic health record for education or training purposes</td>
</tr>
<tr>
<td>Hardware-specific evaluations</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Development phases</th>
</tr>
</thead>
<tbody>
<tr>
<td>No restriction was made with regard to the development phase</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>No setting was excluded</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Participants in development process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exclusively patients or other users</td>
</tr>
<tr>
<td>Trainees or students in the health care sector without patient contacts</td>
</tr>
</tbody>
</table>

Information Sources

The search was carried out in the PubMed, CINAHL, and Scopus databases. A supplementary search was carried out in Google Scholar. The final search was performed by JL and SK on March 17, 2021, using the search strings from Multimedia Appendix 2. The forward-and-backward citation tracking [123] was then carried out by JL using Scopus and Google Scholar.

Search Strategy

First, an initial limited search was conducted in a selection of relevant databases to analyze possible terms in the title and abstract to identify keywords describing the articles. This was followed by a search of all databases using all identified keywords. SK and JL formulated the basic idea of the review and conducted the initial searches. Afterward, SK, JL, and TSB developed the search terms. The search terms used were based on two main categories: (1) search terms around the term EHRs and corresponding synonyms as well as Medical Subject Headings terms (PubMed) and subject headings (CINAHL) were used, and (2) search terms around the term participatory design with corresponding synonyms and Medical Subject Headings terms or subject headings were used. The search strings for PubMed, CINAHL, Google Scholar, and Scopus are shown in Multimedia Appendix 2.
The search in Google Scholar used a substantially shortened search string, as the search engine cannot process longer, complex search strings. This resulted in several results that did not meet the inclusion criteria. Therefore, using Google Scholar’s sort by relevance function, only the first 250 results were checked for eligibility, of which 47 were selected.

Selection Process
All citations were imported into the bibliographic manager EndNote (Clarivate), and duplicate citations were automatically removed, with further duplicates removed if found later in the process. The citations were then imported into the software [124] to subsequently check the relevance of the titles and summaries and to characterize the data of the full articles. Rayyan provided blinded checking and automatically displayed matching inclusions, exclusions, and conflicts after blinding was turned off.

First, the titles and abstracts were checked by SK and JL to ensure compliance with the inclusion criteria. Differences were discussed with TSB. Subsequently, TSB and JL screened titles and abstracts for the forward-and-backward citation tracking results, and the differences were discussed with CJ.

All citations deemed relevant after title and abstract screening were obtained for subsequent review of the full-text article. For articles that could not be obtained through institutional holdings that were available to the authors, attempts were made to contact the authors of the source and request the article. In addition, articles were requested via interlibrary loan.

SK and JL screened the full texts; the differences were discussed with TSB. TSB and JL screened the full texts of the forward-and-backward citation tracking results; the differences were discussed with CJ.

At this stage, studies were excluded if they did not meet the eligibility criteria. After reviewing approximately 25 articles independently, the reviewers met to resolve any conflicts and to ensure consistency among the reviewers and with the research question and purpose [125]. The excluded studies were appropriately labeled with the reason for exclusion to improve traceability.

Data Analysis
Categories were formed deductively. This was based on a systematic review by Vandekerckhove et al [33] that focused on electronic health interventions. The aim of the review by Vandekerckhove et al [33] was to report and justify participatory design methods in empirical eHealth studies for further development of the methodology. The decision to follow this review was based on its comprehensive presentation and its fit for the research question pursued here. However, the categories were supplemented by inductive categories that emerged from reviewing the material. The categories can be named as “factual categories” according to Kuckartz [126], designating specific facts in the included studies. All codes were reviewed, coded, and discussed in regular meetings by TSB, CJ, and SK.

Categories for Syntheses
Owing to the diversity of study designs and the research questions, a quality assessment was not performed. Following the approach of Vandekerckhove et al [33], an assessment of the sufficiency and design of reporting was conducted. To improve comprehensibility, the inductive categories were supplemented by key questions (based on the definitions of the categories that were created and constantly refined during the analysis process) and served to represent the collected data items. This was partly based on the categories in the study by Vandekerckhove et al [33], whose review dealt with eHealth interventions. For example, category 1 in this review was developed based on the category “eHealth intervention” by Vandekerckhove et al [33] and category 3—study participants—was based on the category “stakeholder types” by Vandekerckhove et al [33]. Category 4—methods and materials—was based on the category “tools of participatory design” by Vandekerckhove et al [33]. The other categories were derived from the material itself, as described earlier in the Data Analysis section. This resulted in the following division, which was used to structure the method representation:

1. Focus and scope of the studies: What is mentioned about the characteristics of the EHR to be developed and the stage of the technology (prototype, already implemented EHR)? What was the aim of the studies?
2. Setting: Where did the involvement in the development of the EHR take place?
3. Study participants: Who was involved in the development? Which characteristics were mentioned when describing the study participants?
4. Methods and materials: Which study design was used? Which terminology was used to describe the involvement process? Which methods were used? Are there any physical materials used in the process? How often were participants involved in the process (involvement counts as renewed involvement if it takes place at a later point in time and contributes to the further development of the technology)?
5. Frameworks, theories, and guidelines: What approaches have been used and influence the choice of methods? Which approaches were used only within the data analysis or a specific method without influencing the choice of methods for the entire study? Which design guidelines were mentioned that influenced the basic logic of the EHR design?
6. Competencies of the researchers: What competencies do researchers contribute in terms of development?

Results

Study Selection
The study selection is described in Figure 1.

The initial search resulted in a total of 23,446 hits (PubMed: n=8281, 35.32%; CINAHL: n=1846, 7.87%; Scopus: n=13,319, 56.81%). In addition, 47 records were extracted from other sources (Google Scholar) after screening the first 25 pages of approximately 7710 results. From a total of 23,446 hits from the initial search and these 47 additional records, 19,002 (81.04%) hits remained after duplicate reduction.
Of these 19,002 articles, 18,830 (99.09%) articles were excluded during title and abstract screening. The remaining 172 texts were subjected to full-text screening, resulting in a total of 74 titles.

The forward-and-backward citation search yielded a total of 2769 hits (755 by forward citation and 1985 by backward citation). Automatic deduplication reduced the number of hits to 2665. Manual duplicate reduction led to a final result of 2625 hits. After title and abstract screening (34 texts remaining), full-text screening was carried out, resulting in 23 articles.

These 23 articles from the forward-and-backward citation search were included in the final assessment along with the previous 74 studies from the initial search. Of these 97 studies, 27 (28%) studies were excluded because of insufficient information and duplicates. The remaining 70 articles were included in the evaluation and can be found in the metadata table in Multimedia Appendix 1.

Figure 1. Flow diagram of the study selection process. EHR: electronic health record.

Study Characteristics
Multimedia Appendix 1 includes the metadata of the included studies.

Results of Syntheses
Focus and Scope of the Studies
The included studies targeted different objectives in EHR development, which can be divided into 8 groups. Some of the studies focused primarily on eliciting users' needs and wants toward EHRs. This included (1) studies to collect general information on user needs and preferences [47-50,63,84,88,93,104-106,112] or (2) studies focusing on factors for implementation [60,69,85,111]. Other studies were oriented toward actual implementation and (3) described pilot-testing [90] or (4) the overall design process [65,76,77,114]. Further studies were based on refining the existing content such as (5) studies that focused on the redesign of EHRs or prototypes [57,62,66] or (6) studies that included system improvement and further development [57,91,95,99]. In addition, studies have examined implemented EHRs, which were (7) studies in terms of overall satisfaction or acceptance [51,64,67,68,79-81,86,96,107] or (8) studies focusing on terms of usability or system performance [52-54,58,59,61,70-75,78,87,89,92,94,97,98,100-103].

The included studies covered different forms of EHRs and development, which were divided into 4 different categories:

1. Information needs for subsequent programming of EHRs were the area of research of 4 studies [47-50]. In these studies, the authors addressed the information needs in EHRs from the perspective of clinicians. For example, Acharya et al [47,51] gathered information needs for oral health information for an EHR, and Ellsworth et al [49]...
surveyed the information needs for a neonatal intensive care EHR.

2. In total, 5 studies reported on prototypes or mock-ups [52-56]. For example, Belden et al [57] addressed a clinical note prototype, and Horsky et al [54] developed a prototype of an EHR that allowed clinicians to complete a summary for outpatient visits.

3. Entire EHRs, including all submodules were addressed in 30 studies [56, 58-86]. For example, Nolan et al [73] examined information use and workflow patterns for an EHR in an ICU.

4. Individual modules of an EHR were addressed in 31 studies [51, 57, 70, 87-113]. These studies focused on individual modules of an EHR rather than an entire EHR. In this category, for example, Aakre et al [87] focused on a module for the automatic calculation of a sequential organ failure assessment calculator in sepsis detection, and Ahluwalia et al [88] focused on dyspnea assessment for palliative care.

**Setting**

The included studies were conducted in the context of different health care settings. A total of 25 studies were conducted in an unspecified hospital setting [60, 62, 65, 67-72, 75-78, 80-85, 89, 91, 92, 99, 100, 111, 114]. Furthermore, 8 studies were conducted in the context of ICUs [49, 55, 56, 63, 68, 73, 87, 93, 115], 6 were conducted in the context of family medicine [52, 59, 101, 102, 109, 110], 6 were conducted in primary care hospitals [53, 61, 97, 104, 105], 5 were conducted in outpatient or clinic settings [54, 57, 74, 95, 108], and 3 were conducted in tertiary hospitals [50, 86, 107]. Two studies each were conducted in a dental clinic [47, 51], in a palliative care setting [88, 96], in emergency departments [90, 103], in the gynecological and antenatal settings [64, 66], and in the context of mental health or psychiatry settings [48, 69]. One study each was conducted in the setting of home care [94], older adult care [79], community health [106], cancer centers [112], and childcare [113].

**Study Participants**

The participants in the included studies comprised a total of 15 different professions (Table 1). Physicians were involved in 76% (53/70) of the studies, whereas nurses were involved in 40% (28/70) of the studies. Pharmacists were involved in 10% (7/70) of the studies, physiotherapists were involved in 6% (4/70) of the studies, social workers were involved in 4% (3/70) of the studies, and medical assistants were involved in 3% (270) of the studies. In 16% (11/70) of the studies, the user groups were not specified. Demographic characteristics such as age, sex, and education were described in 21 studies [53, 57, 67-70, 73, 75, 79, 81, 84, 91-93, 96, 98, 103, 105, 106, 111, 113, 114]. Moreover, in 3 studies, the authors provided a brief description of demographic characteristics [64, 74, 97]. For example, one study provided a description of the demographic characteristics as follows: “The sample consisted of 21 female participants and 9 male participants, with a proportion of 70% female and 30% male” [64]. The remaining studies did not describe the demographic characteristics of the participants.

In 10% (7/70) of the included studies, participants received financial compensation for taking part in the study. The following amounts were paid to the participants: US $100 gift card—25 to 45 minutes [92], US $100 gift card [68], US $50 per hour [75], US $100 per 2 hours [76], €40 (US $43) per hour [96], and US $100 per hour [56]. In one study, a US $25 gift card for a restaurant was offered [59].

In addition to the characteristics of the participants, the number of participants in each study was divided into 7 categories (Table 2).

**Table 1. User groups included in the studies (n=70).**

<table>
<thead>
<tr>
<th>User group</th>
<th>Studies that included this user group, n (%)</th>
<th>Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physicians</td>
<td>52 (74)</td>
<td>[48-50, 52-60, 62-71, 73, 75, 76, 78, 79, 82, 83, 86, 88, 89, 91-93, 95-97, 100-105, 107-115]</td>
</tr>
<tr>
<td>Nurses</td>
<td>29 (41)</td>
<td>[48-50, 60, 64, 65, 69, 72, 74, 78, 81, 84, 85, 88, 93, 94, 96, 99, 100, 102, 103, 105-107, 111, 113]</td>
</tr>
<tr>
<td>Pharmacists</td>
<td>7 (10)</td>
<td>[65, 78, 79, 82, 85, 90, 91]</td>
</tr>
<tr>
<td>Physiotherapists</td>
<td>4 (6)</td>
<td>[60, 78, 79, 97]</td>
</tr>
<tr>
<td>Social workers</td>
<td>3 (4)</td>
<td>[79, 85, 93]</td>
</tr>
<tr>
<td>Medical assistants</td>
<td>2 (3)</td>
<td>[74, 106]</td>
</tr>
<tr>
<td>Psychologists</td>
<td>1 (1)</td>
<td>[85]</td>
</tr>
<tr>
<td>Physician assistants</td>
<td>1 (1)</td>
<td>[106]</td>
</tr>
<tr>
<td>Managers</td>
<td>1 (1)</td>
<td>[85]</td>
</tr>
<tr>
<td>Medical office assistants</td>
<td>1 (1)</td>
<td>[98]</td>
</tr>
<tr>
<td>Midwives</td>
<td>1 (1)</td>
<td>[66]</td>
</tr>
<tr>
<td>Community health agents (CHA)</td>
<td>1 (1)</td>
<td>[48]</td>
</tr>
<tr>
<td>Primary care providers</td>
<td>1 (1)</td>
<td>[92]</td>
</tr>
<tr>
<td>Medical secretaries</td>
<td>1 (1)</td>
<td>[60]</td>
</tr>
<tr>
<td>IT departments, hospital’s IT</td>
<td>1 (1)</td>
<td>[60]</td>
</tr>
<tr>
<td>Not specified</td>
<td>11 (16)</td>
<td>[47, 51, 57, 61, 62, 77, 78, 85, 87, 105, 113]</td>
</tr>
</tbody>
</table>
Table 2. Number of participants per study (n=70).

<table>
<thead>
<tr>
<th>Range for the number of participants</th>
<th>Studies, n (%)</th>
<th>Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-10</td>
<td>15 (21)</td>
<td>[54-56,63,77,82,83,89,95,97,98,100,108,109,115]</td>
</tr>
<tr>
<td>11-20</td>
<td>19 (27)</td>
<td>[52,53,58-61,66,74-76,87,88,92,96,99,103,104,110]</td>
</tr>
<tr>
<td>21-30</td>
<td>12 (17)</td>
<td>[49,57,64-67,69,72,73,90,102,107,114]</td>
</tr>
<tr>
<td>31-40</td>
<td>5 (7)</td>
<td>[71,79,84,94,105]</td>
</tr>
<tr>
<td>41-50</td>
<td>5 (7)</td>
<td>[48,65,80,91,101]</td>
</tr>
<tr>
<td>51-100</td>
<td>5 (7)</td>
<td>[51,78,106,111,113]</td>
</tr>
<tr>
<td>&gt;100</td>
<td>9 (13)</td>
<td>[47,50,62,70,81,85,86,93,112]</td>
</tr>
</tbody>
</table>

**Methods and Materials**

First, the basic methodology of the studies was examined. Overall, 36% (25/70) of the studies used mixed methods design [54,55,57,60,61,66,72,75,78,79,81,83,84,90,92,94,95,98,100,102,103,105,108,109,111,113]. Moreover, 31% (22/70) of the studies used a qualitative design [49,50,52,56,59,62,64,69-71,73,74,77,86,87,91,96,99,104,107,114,115], whereas 31% (22/70) of the studies used a quantitative design [47,48,51,53,58,63,65,67,68,76,80,82,85,88,93,97,99,101,106,110,112].

The wide variance of terminology in relation to the involvement of users in technology development already mentioned at the beginning is also reflected in the articles included. The terminology here describes the literal naming of the method by the authors of the studies themselves, regardless of how it was conducted. The largest proportion of studies (44/70, 63%) did not include a designation of methodology [47,51,52,55-62,64,66,67,69,71,72,76-86,88,90,93,95,98-101,104,107,110,111,115]. For example, in 23% (16/70) of the studies, the authors of the respective manuscripts described the methodological approach as “user-centered design” [54,57,68,70,73-75,79,87,93,97,102,103,108,109,114]. In one of the studies, it was only given as a keyword and not in the manuscript [97]. In terms of frequency, the following terms were used: “participatory design” in 6% (4/70) of the studies [48,63,65,96], “co-design” in 3% (2/70) of the studies [94,105], and “iterative rapid design involving providers” [53], “end-user design” [49], “multidisciplinary design” [61] and “human-centered design” [113] in 1 study each.

The frequency with which users were involved in the development was examined. Involvement counts as renewed involvement if it takes place at a later point and contributes to the further development of the technology (for example, several surveys for the iterative refinement of a prototype). In 57% (40/70) of the studies, users were included once [47,49-51,55,56,58,59,69-73,75-77,80,82-84,88,91-93,95-97,99-101,103-106,108-111,114,115]. An involvement of users at 2 points was investigated in 24% (17/70) of the studies. Moreover, 9% (6/70) of the studies reported 3 times of user involvement, 4% (3/70) of the studies reported 4 times of involvement, 23% (2/70) of the studies reported 6 times of involvement, 1% (1/69) study reported 5 times of involvement, and 1% (1/69) study reported 9 times of involvement.

In some of the studies, a foundation of the study was provided before the actual (further) development of the EHR. This included, for example, literature reviews [47,61,63,68,90], pilot-testing of the design [52], pilot-testing of the survey [81] or interview guide [47,51,68], a review of 12 different EHRs [57] as well as training with the software in advance [76,77,83,103,107], and the presentation of learning videos [91].

A common method of data collection and involvement of health care professionals was to test a prototype as a walkthrough using think-aloud technique [52,55,56,58,59,71,74-77,83,89,90,92,93,95-100,103-105,108-110,113-115]. As part of the walkthrough methodology, various programs (eg, Morae) have been used to record audio or screen displays, mouse clicks, and keyboard [52,54,74-76,83,92,94,97,100,103,110,113,114].

A related method, the near-live testing, was used in one study [89].

Another common method were the questionnaires. In addition to individually created surveys [47,49,50,60,64,66,70,78,79,83,86,101-103,106,107,113], various existing questionnaires were used (Table 3).

Some of the studies used web-based questionnaire tools (eg, Survey Monkey) [47,50,61,69,70,86,113].

Individual semistructured interviews [48,53-55,60,63,65,67,68,70,72,79,80,82-85,87,88,90,93,94,96-98,102,104-106,110,111,113,114] and group interviews and focus group discussions [48,60,62,63,67] were conducted. In some of the studies, design workshops were held with various users [53,57,65,90,110].

One method that was often combined with interviews was observation. This involved observing health care professionals as they used an EHR to conduct documentation [60,63,66,73,79,85,87,90,94,102,113,114]. This includes observations in both a clinical and a study setting.

The use of mock-ups was another common method in the studies. This contained paper prototypes [57,90,94] and web-based prototypes using different prototyping tools (eg, HipMunk) [57,87,90,93,94,105,113,114].
Table 3. Questionnaires used in the studies.

<table>
<thead>
<tr>
<th>Questionnaire</th>
<th>Focus of the questionnaire</th>
<th>Studies using the questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baylor EHR(^a) user experience survey</td>
<td>Measuring user experience following EHR implementation</td>
<td>[69]</td>
</tr>
<tr>
<td>Canada Health Infoway System and Use Assessment Survey</td>
<td>Measuring user adoption and use as well as information and system quality</td>
<td>[81]</td>
</tr>
<tr>
<td>Computer Systems Usability Questionnaire</td>
<td>Measuring satisfaction of users with computer system usability</td>
<td>[87]</td>
</tr>
<tr>
<td>Information System Use Instrument</td>
<td>Measuring nurses’ information systems use</td>
<td>[81]</td>
</tr>
<tr>
<td>Keystroke-level model GOMS</td>
<td>Predict or estimate the time for completing a task in software</td>
<td>[59]</td>
</tr>
<tr>
<td>Nasa Task Load Index</td>
<td>Measuring perceived workload</td>
<td>[52,58,59,92,113]</td>
</tr>
<tr>
<td>Physician Documentation Quality Instrument-9</td>
<td>Assessing the quality of physician electronic documentation</td>
<td>[75]</td>
</tr>
<tr>
<td>Post-Study System Usability Questionnaire</td>
<td>Measuring the perceived satisfaction</td>
<td>[100]</td>
</tr>
<tr>
<td>Questionnaire for user interaction satisfaction (short form)</td>
<td>Measuring the subjective satisfaction with the human-computer interface</td>
<td>[94]</td>
</tr>
<tr>
<td>Single Ease Question</td>
<td>Assessing the difficulty of a task</td>
<td>[74,90]</td>
</tr>
<tr>
<td>System Usability Scale</td>
<td>Measuring the usability</td>
<td>[52,55,70,72,75,77,78,94,100,103,108]</td>
</tr>
<tr>
<td>Technology Acceptance Model Questionnaire</td>
<td>Measuring likelihood of technology acceptance</td>
<td>[61]</td>
</tr>
<tr>
<td>Usability Assessment</td>
<td>Measuring usability</td>
<td>[72]</td>
</tr>
<tr>
<td>Workflow Integration Survey</td>
<td>Measuring workflow integration</td>
<td>[81]</td>
</tr>
</tbody>
</table>

\(^a\)EHR: electronic health record.

Less frequently used methods include document analysis [48,85] and extraction of routine data from the EHR for analysis [60,61,70,74,111,113].

An overview of the respective methods by study aim is available in the metadata table in Multimedia Appendix 1.

### Frameworks, Theories, and Guidelines

In the following studies, frameworks are understood as approaches that frame the entire research project and influence the choice of methods and their structure. In these studies, the updated DeLone and McLean framework [127] for evaluating information systems success was used once [60]. The design science framework [128] was used once to develop prototype dashboards [94]. In the same study, the tasks, users, representations, and functions framework [129] was used to structure the usability evaluation [94]. Falah et al [64] used the plan-do-study-act cycle [130] to facilitate the implementation process. In addition, Dziadzko et al [86] used the define-measure-analyze-improve-control quality measurement [131] for implementation measurement. The social science approach of lightweight ethnography [132] was used to design the study by Chruscicki et al [90]. Owing to limited time, the predesigned sample, and particular research questions, one study [67] used framework analysis [133] as a framework. Sockolow et al [79] used the health information technology research-based evaluation framework [134] as a framework to design their study and as a theory to merge qualitative and quantitative data. In addition, in 7% (5/70) of the studies, the authors referred to ISO 9241-210 in their theoretical background [59,70,75,113,114]. However, ISO 9241-210 was not used as a theoretical framework in any of the incident studies.

Theories are understood to be those approaches that were used exclusively within the data evaluation or specific analytic method but did not influence the choice of methods for the entire study as a whole. Kernebeck et al [96] used the Unified Theory of Acceptance and Use of Technology [135] to evaluate think-aloud sessions. Wawryniak et al [82] used the critical incident technique by Flanagan [136] to design interview protocols. The interview guide in another study [105] was based on the diffusion of innovations theory [137] and complementary ones. The human factors model Systems Engineering Initiative for Patient Safety 2.0 [138] was used by Cohen et al [106] for data collection. The Cognitive Load Theory [139] was described as important and used for evaluation by Curran et al [92]. In addition, the attention capacity model [140], which focuses on mental effort, was used as a theoretical background by Mosaly et al [71]. The technology acceptance model [141] was used to design a questionnaire [113].

In addition, design guidelines were mentioned that influenced the basic logic of the EHRs design. This includes the ergonomics of activity [142] mentioned in one study [48], the suggested time and motion procedures [143] in another study [73], and usability heuristics by Nielsen [144] in 1 study [74]. Another framework used was the spiral model for software development [145] in addition to the EHR system user interface framework of the Veterans Affairs Computerized Patient Record System [146] in 1 study [95]. The data-knowledge-information-wisdom...
Competencies of Researchers

In addition, all studies were screened for the description of the competencies of the researchers who conducted the studies. A distinction can be made between competences related to software knowledge (e.g., usability experience and software programming) or competences related to knowledge of the context of use (e.g., previous experience of working with EHRs in clinical settings) or methodological skills in the area of data collection (qualitative interviews, surveys). In 30% (21/70) of the studies, the authors briefly described the competencies of the researchers [49,52,53,60,63,68,70,74,75,82,88,93,97,99,103,104,106,109,110,112,113]. Examples of these descriptions were that the researchers described themselves as “experienced in qualitative research” [106] and “the research team included three academic researchers and two clinical nurses” [99].

Discussion

This review aimed to provide an overview of the existing methods of user involvement in the literature for developing and evaluating EHRs.

Principal Findings

The review had four objectives: (1) to conduct a systematic search of the published literature for studies focusing on user involvement in the development of EHRs, (2) to present the characteristics and range of methods used in the identified manuscripts, (3) to explore the reported challenges and limitations of the methods, and (4) to make recommendations for further developing the approach and improving the consistency with which they are conducted and reported. Therefore, the main focus of the review was to examine in which settings which participants were involved with which methods and materials and which frameworks were used. Furthermore, the frequency and design of the development and an overview of the competences of the respective researchers involved in the development were examined. To the best of our knowledge, this is the first review to describe the methodological aspects for involving health care professionals in the development of EHRs.

The characteristics of EHRs addressed in the included studies covered a variety of different aspects. On the one hand, a large number of studies addressed a comprehensive EHR, whereas on the other hand, many studies addressed only individual modules of an EHR. The wide range of characteristics in this review was largely because of the broad inclusion criteria, which were designed to provide a comprehensive picture of the methodological aspects of professional involvement in the development of EHRs. This leads to a better description of the complex field of EHRs and their methodological aspects. However, this sometimes makes it difficult to compare the interventions. In terms of setting, it was found that most of the included studies were conducted in general hospitals and ICUs. The authors suggest analyzing studies in a specific setting and how the participants are involved there in the future to cover the methodological aspects, such as in ICU or in palliative care.

Future studies could also focus on the individual modules of an EHR, for example, medication modules.

In terms of participants, the studies mainly included physicians and nurses. In terms of multidisciplinary care, it would be desirable for all health care professionals (including physiotherapists, occupational therapists, speech therapists, social workers, nursing assistants, etc) to record their activities and observations in joint documentation and to be able to view them mutually. In addition, sharing EHRs can facilitate communication [148] (e.g., by sending messages within a program and assigning tasks). With this in mind, it is surprising that only these 2 professional groups were so intensively involved. It would be desirable for further studies to include all professional groups and to design EHRs to meet their needs. However, most of the studies did not specify which health care professionals were involved in the development. This was partly because of an imprecise naming of the participants and partly because of the lack of naming. Future studies should specifically describe the demographic characteristics of the participants, which may lead to a better assessment of the results [33,36]. This is important because demographic variables have a strong influence on the acceptance of EHRs and the level of competence in using EHRs and digital health technologies in general [149,150]. Therefore, it is recommended that study investigators collect key demographic variables from participants and present them in tabular form to improve the interpretation of the results.

The methodology of the studies was balanced between qualitative, quantitative, and mixed methods. However, in 63% (44/70) of the studies, no terminology was used to describe the design of user involvement in more detail. This adds to the imprecision of the presentation in terms of the level of involvement and a qualitative assessment of the methodology. Although 23% (16/70) of the included studies that mentioned user-centered design as an approach will be examined to see if this was really implemented, most of the studies remain vague about user involvement. This again supports the broad search strategy of the review but also points to qualitative ambiguities of implementation.

The frequency of user involvement varied widely, and in most of the studies (40/70, 57%), users were involved only at one point. This shows that a true participatory design or co-design, as it is called for, is rather rare and fuels the suspicion of sham participation, where user requirements are collected but no iteration is performed to test the fit. Another problem with user involvement in development is that it often occurs at only one point in the development of new technologies. This problem is often referred to as “project-based temporality” in the involvement of users [151]. Therefore, it is recommended that users are involved in the development of new technologies at all stages of development over a longer period [151].

Most of the methods vary widely. Think-aloud approaches were often used to obtain user feedback on an EHR. The continuation of this method, near-live testing, which also increases the likelihood of a good fit, was used in only one study. In the case of questionnaires, individual, nonvalidated questionnaires were frequently used, which reduced the quality of the results. The
most commonly used questionnaires were mainly oriented toward usability (System Usability Scale) in 16% (11/70) of the studies and the cognitive load (Nasa Task Load Index) in 7% (5/70) of the studies. Cognitive load refers to the amount of mental effort required by a person to perform a specific task and the associated required capacity of the human working memory [152]. It is imperative to consider cognitive load in the development of EHRs and in the implementation of appropriate solutions in clinical practice, as cognitive load has been linked to the development of burnout and distress [27]. It is recommended that user involvement studies use a mix of methods from the fields of telling, making, and enacting [153]. This emphasizes the impact of user involvement through the exchange of current and future practices and the sharing of needs (telling). For successful user involvement, it is crucial that future users develop something (making), contributing to the existence and design of a new technology. In addition, by involving users, it should be possible to transfer ideas into reality by creating a simulation to test them [153]. Accordingly, different questions require different methods from each field, but a mixed methods approach ensures a diversity of perspectives.

Frameworks, theories, and guidelines were very rarely used. Moreover, the results showed a rather low level of consideration of the theoretical underpinnings to the detriment of the quality of the studies. This is particularly problematic because the use of such frameworks can structure the development and make the replicability of results between different studies comparable. Furthermore, it is problematic that a large number of studies did not refer to theories and models. This would also provide a theoretical basis for the development and make the quality of the results more comprehensible [32]. It is particularly surprising that only 9% (6/70) of the studies referenced to ISO standards. These standardizes the process for the development of new software. It would be useful to refer to these standards in future studies and to highlight the stage of development of the respective technology [154,155].

The researchers’ skills have rarely been documented. A multidisciplinary research and development team should consist of individuals with different skills from different health care disciplines, methodological disciplines, and social disciplines. This allows for optimal design and support of different stakeholders during development and implementation [156]. Similar to the sometimes imprecise description of demographic characteristics of the study participants, the skills of the investigators should be described. Owing to the interdisciplinary nature of technology development research projects, it would be advisable to have multidisciplinary teams and to identify the respective competencies and experience in technology development.

**Limitations**

To be able to interpret these results, it is necessary to describe several limitations of this study. First, the search strategy was limited by its focus on empirical, scientifically published work. The involvement of health care professionals in the development of EHRs may not always be published in scientific journals. Therefore, this review is a first step on the topic of involving health care professionals in the development of EHRs. In further studies, it might be interesting to include gray literature and databases with a focus on technology-oriented research and engineering (eg, Institute of Electrical and Electronics Engineers). However, the heterogeneity of the quality of the publications must be taken into account. In addition, EHRs are mostly developed by large digital technology companies. It can be assumed that these companies often involve users in the development but do not produce publications or perform actual research. Therefore, it can be assumed that there was publication bias. It would be necessary in the future to survey such large companies on how they involve users in the development of EHRs.

Second, it should be noted that the screening process was limited by the definitions of user involvement, which accordingly shaped both the search terms and the inclusion and exclusion criteria. Although a wide range of terms were used, it cannot be ruled out that individual manuscripts that were coherent in terms of content, and therefore would have led to different results, were not included because of the lack of used terminology.

Third, it should be taken into account that some studies published their results in several manuscripts and did not briefly review the entire development or implementation process. In our evaluation, we were only able to consider the described frequency of user involvement from the information provided in the included manuscripts. However, it is possible that the included manuscripts each report only a subset of the study project; whereas in the overall study project, users were much more frequently involved.

Finally, one of the findings, namely the sometimes-low transparency of reporting, also directly points to a limitation in terms of analysis and conclusions—drawing conclusions on the basis of reporting in manuscripts has limited validity, as there is no way to ensure that the actual methodological considerations and intentions correspond to what was presented in the manuscript.

**Further Research**

Further research can help to improve the methodological framework for involving health care professionals in the development of EHRs. Particular attention should be paid to the rationale for the methodological choices. It is also crucial to combine different methods from the fields of telling, making, and enacting; involve users at several points in time to avoid sham participation; and strive for maximum user orientation. The growing interest in “design through design research” [157] should be encouraged but with conditions to promote high-quality developments. Little knowledge can be gained from publications with low reporting quality in terms of transferability and quality assessments. It would be useful for studies to report the aspects more precisely. Specific reporting guidelines for reporting the results of technology development studies would be helpful, as is the case for many other types of studies [158]. Process evaluations should be used in a standardized manner to improve study quality.

In addition, in future studies, it would be necessary to examine in more detail the outcomes of the participatory design with
users. In this context, questions should be answered regarding the specific outcomes that have been improved by the involvement of users. How these outcomes were measured and how, for example, improvements to the software were evaluated in different iterations should also be analyzed.

Conclusions
Studies involving health care professionals in the development of EHRs have used various approaches. This paper provides an overview of the approaches in different fields of development with the inclusion of diverse users. Often, however, there is no specific approach, framework, or theory underlying the procedure and there is missing or inaccurate information in the reporting.

Acknowledgments
The authors acknowledge support from the Open Access Publication Funds of the Ruhr-Universität Bochum.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Metadata of the included studies.
[DOCX File, 65 KB - humanfactors_v10i1e45598_app1.docx]

Multimedia Appendix 2
Search string.
[DOCX File, 27 KB - humanfactors_v10i1e45598_app2.docx]

References
Merkel S, Kucharski A. Participatory design in gerontechnology: a systematic literature review. Gerontologist 2019 Jan

Fischer B, Peine A, Östlund B. The importance of user involvement: a systematic review of involving older users in
technology design. Gerontologist 2020 Sep 15;60(7):e513-e523 [FREE Full text] [doi: 10.1093/geroni/gnx004.1423]

Panek I, Crumley ET, Ishigami-Doyle Y, Sixsmith J, Kontos P, O’Doherty K, et al. Levels of older adults’ engagement in
technology research, design and development: a scoping review. InnoV Aging 2017 Jul;1(suppl_1):393 [FREE Full text]
[doi: 10.1093/geroni/ignx033]

Clemensen J, Rothmann MJ, Smith AC, Caffery LJ, Danbjorg DB. Participatory design methods in telemedicine research.

Joddrell P, Astell AJ. Studies involving people with dementia and touchscreen technology: a literature review. JMIR Rehab
Assist Technol 2016 Nov 04;3(2):e10 [FREE Full text] [doi: 10.2196/rehab.5788] [Medline: 28582254]

DeSmet A, Thompson D, Baranowski T, Palmeira A, Verloigne M, De Bourdeaudhuij I. Is participatory design associated
Apr 29;18(4):e94 [FREE Full text] [doi: 10.2196/jmir.4444] [Medline: 27129447]

Göttgens I, Oertelt-Prigione S. The application of human-centered participatory approaches in health research and innovation: a
narrative review of current practices. JMIR Mhealth Uhealth 2021 Dec 06;9(12):e28102 [FREE Full text] [doi: 10.2196/28102]
[Medline: 34874893]

2013.

Vargas C, Whelan J, Brimblecombe J, Allender S. Co-creation, co-design, co-production for public health - a perspective
[Medline: 35702744]

Acharya A, Mahnke A, Chyow PH, Rottscheit C, Starren JB. Medical providers’ dental information needs: a baseline survey.

Acharya A, Shimpni N, Mahnke A, Mathias R, Ye Z. Medical care providers’ perspectives on dental information needs in


Misharis RG, Yoder J, Wilson D, Mann D. Integrating data from an online diabetes prevention program into an electronic
health record and clinical workflow, a design phase usability study. BMC Med Inform Decis Mak 2016 Jul 11:16:88 [FREE

Horsky J, Ramelson HZ. Development of a cognitive framework of patient record summary review in the formative phase of

King AJ, Cooper GF, Hochheiser H, Clermont G, Visweswaran S. Development and preliminary evaluation of a prototype
[Medline: 26958296]

King AJ, Cooper GF, Hochheiser H, Clermont G, Hauskrecht M, Visweswaran S. Using machine learning to predict the
information seeking behavior of clinicians using an electronic medical record system. AMIA Annu Symp Proc 2018 Dec
05;2018:673-682 [FREE Full text] [Medline: 30815109]


72. Nation J, Wangia-Anderson V. Applying the data-knowledge-information-wisdom framework to a usability evaluation of electronic health record system for nursing professionals. Online J Nurs Inform 2019;23(1) [FREE Full text]


123. Busse et alJMIR HUMAN FACTORS


145. Boehm BW. A spiral model of software development and enhancement. Computer 1988 May;21(5):61-72 [FREE Full text] [doi: 10.1109/2.59]


Abbreviations

EHR: electronic health record
ICU: intensive care unit
ISO: International Organization for Standardization
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

©Theresa Sophie Busse, Chantal Jux, Johannes Laser, Peter Rasche, Horst Christian Vollmar, Jan P Ehlers, Sven Kernebeck. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 10.07.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits

https://humanfactors.jmir.org/2023/1/e45598

JMIR Hum Factors 2023 | vol. 10 | e45598 | p.1068
An Artificial Therapist (Manage Your Life Online) to Support the Mental Health of Youth: Co-Design and Case Series

Aimee-Rose Wrightson-Hester1,2,3, BA; Georgia Anderson4; Joel Dunstan5; Peter M McEvoy1,2,6, PhD; Christopher J Sutton7, PhD; Bronwyn Myers1,8,9, PhD; Sarah Egan1,2, PhD; Sara Tai10, DClinPsy; Melanie Johnston-Hollitt5, PhD; Wai Chen1,11,12,13, PhD; Tom Gedeon14, PhD; Warren Mansell1,2,10, DClinPsy

1Curtin enAble Institute, Faculty of Health Sciences, Curtin University, Perth, Australia
2Discipline of Psychology, School of Population Health, Curtin University, Perth, Australia
3School of Arts and Humanities, Edith Cowan University, Perth, Australia
4Mental Health Commission of Western Australia, Perth, Australia
5Curtin Institute for Data Science, Curtin University, Perth, Australia
6Centre for Clinical Interventions, North Metropolitan Health Service, Nedlands, Australia
7Centre for Biostatistics, School of Health Sciences, The University of Manchester, Manchester, United Kingdom
8Alcohol, Tobacco and Other Drug Research Unit, South African Medical Research Council, Parow, South Africa
9Division of Addiction Psychiatry, Department of Psychiatry and Mental Health, University of Cape Town, Cape Town, South Africa
10Department of Clinical Psychology, School of Health Sciences, The University of Manchester, Manchester, United Kingdom
11Mental Health Service, Fiona Stanley Hospital, Perth, Australia
12Curtin Medical School, Curtin University, Perth, Australia
13Centre of Excellence in Medical Biotechnology, Faculty of Medical Science, Naresuan University, Phitsanulok, Thailand
14Optus-Curtin Centre of Excellence in AI, School of Electronic Engineering, Computing and Mathematical Sciences, Curtin University, Perth, Australia

Corresponding Author:
Aimee-Rose Wrightson-Hester, BA
Curtin enAble Institute
Faculty of Health Sciences
Curtin University
Kent St
Perth, WA 6102
Australia
Phone: 61 452491337
Email: aimee.wrightson-hester@curtin.edu.au

Abstract

Background: The prevalence of child and adolescent mental health issues is increasing faster than the number of services available, leading to a shortfall. Mental health chatbots are a highly scalable method to address this gap. Manage Your Life Online (MYLO) is an artificially intelligent chatbot that emulates the method of levels therapy. Method of levels is a therapy that uses curious questioning to support the sustained awareness and exploration of current problems.

Objective: This study aimed to assess the feasibility and acceptability of a co-designed interface for MYLO in young people aged 16 to 24 years with mental health problems.

Methods: An iterative co-design phase occurred over 4 months, in which feedback was elicited from a group of young people (n=7) with lived experiences of mental health issues. This resulted in the development of a progressive web application version of MYLO that could be used on mobile phones. We conducted a case series to assess the feasibility and acceptability of MYLO in 13 young people over 2 weeks. During this time, the participants tested MYLO and completed surveys including clinical outcomes and acceptability measures. We then conducted focus groups and interviews and used thematic analysis to obtain feedback on MYLO and identify recommendations for further improvements.

Results: Most participants were positive about their experience of using MYLO and would recommend MYLO to others. The participants enjoyed the simplicity of the interface, found it easy to use, and rated it as acceptable using the System Usability Scale. Inspection of the use data found evidence that MYLO can learn and adapt its questioning in response to user input.
found a large effect size for the decrease in participants’ problem-related distress and a medium effect size for the increase in their self-reported tendency to resolve goal conflicts (the proposed mechanism of change) in the testing phase. Some patients also experienced a reliable change in their clinical outcome measures over the 2 weeks.

Conclusions: We established the feasibility and acceptability of MYLO. The initial outcomes suggest that MYLO has the potential to support the mental health of young people and help them resolve their own problems. We aim to establish whether the use of MYLO leads to a meaningful reduction in participants’ symptoms of depression and anxiety and whether these are maintained over time by conducting a randomized controlled evaluation trial.

(JMIR Hum Factors 2023;10:e46849) doi:10.2196/46849

KEYWORDS
mental health; conversational agents; chatbots; young people; acceptability; feasibility; co-design; artificial therapist; artificial intelligence; youth; child; adolescent; chatbot; Manage Your Life Online; MYLO; support; mobile phone

Introduction

Background

Globally, the prevalence of child and adolescent mental health issues has increased during the COVID-19 pandemic, up to a current rate of 1 in 5 [1]. Despite this increase, global government spending on mental health services remains low (2%), with shortages of skilled workers and a significant treatment gap between demand and provision for mental health disorders [2,3]. Digital interventions, including mental health-based smartphone apps, that do not require guidance from mental health workers could be one solution for improving timely and equitable access to mental health support worldwide. Therefore, this paper reports the development of a mental health-based smartphone app, Manage Your Life Online (MYLO), and assesses the acceptability and feasibility of this app to support the mental health of young people.

Several reviews have highlighted the benefits of using digital mental health apps (both on the web and offline) to improve consumer access to timely interventions by overcoming many traditional barriers to help seeking and enhancing therapeutic outcomes [4]. Mental health apps may be particularly well placed as a treatment option for adolescents and young people given the high levels of smartphone ownership worldwide [5-7] and initial reviews showing significant improvements in symptoms following app interventions [8]. Although apps provide an opportunity to reach youth who may have limited access to traditional mental health services, it is critical that such digital apps are theory driven, evidence supported, and highly engaging. However, a recent umbrella review (including 36 reviews conducted until 2022) found limited overall empirical and theoretical evidence for the efficacy of these apps or the therapeutic interventions they use [4]. Most apps use strategies based on therapy modalities and lack a theoretical underpinning or use >1 strategy or theory [9]. This makes it difficult to measure and draw conclusions on the most effective modality or theory to use in mental health apps and on how to improve mental health apps. Furthermore, limited user engagement and retention have been a pervasive issue across mental health apps [4,10], and this is largely driven by the user’s dissatisfaction with the functionality of the apps [11].

Conversational agents, or chatbots, that use artificial intelligence technology are a promising and fast-growing subset of mental health apps [12,13] that may be more engaging and therefore have higher levels of self-Adherence than noninteractive apps [14]. Furthermore, as 71% of young people already report using messaging apps with peers to support their mental health, conversational agents can leverage users’ familiarity with texting to provide evidence-based support in a format with which users are already comfortable [15]. However, empirical evidence for the use of chatbots is currently lacking [14,16,17], and many apps are not designed and built according to a robust theoretical basis for a therapeutic paradigm [10,18]. For example, many use an eclectic mix of strategies (such as Tess [19], Wysa [20], and eSmart-MH [25]), and although this may offer users choice within the app, it becomes difficult to draw conclusions on which specific features and strategies are effective or not. Therefore, more research is needed to demonstrate the efficacy of conversational agents, including greater transparency and evaluation of the proposed mechanisms of action used [10,22].

Recent studies focusing on the user experience to identify ways to improve the uptake and engagement of mental health chatbots have generally found high user satisfaction [14,23]. Users have indicated that they value the interactive conversational approach and appear to build a relationship with the chatbots akin to that of a human therapist or friend [10,14]. These findings are consistent across chatbots that use a character or avatar for the agent (eg, Woebot [24], Wysa [20], and eSmart-MH [25]) and those that do not (eg, Tess [19]). Common challenges affecting conversational agents that may impact user engagement and satisfaction include repetitive content, limitations to the agent’s ability to understand the users’ expressed feelings or thoughts, inappropriate response to the user’s statements [10], and usability and technical issues [12].

Another challenge affecting engagement and efficacy of conversational agents, and mental health apps more broadly, is that many apps typically offer disorder-specific interventions rather than transdiagnostic (ie, effective for multiple mental disorders) or universal interventions. Universal interventions and apps use theories and therapeutic techniques that help reduce distress regardless of whether the symptom pattern or severity threshold conforms to those of a formal mental disorder (based on the narrow diagnostic criteria of the Diagnostic and Statistical Manual of Mental Disorders [26] or International Classification of Disease systems [27]) or the precise etiological factors driving the symptoms and impairments [28,29]. A universal approach could lead to increased user engagement and treatment efficacy.
by reducing the burden on users with multiple or overlapping comorbidities by removing the need to use multiple apps. Furthermore, universal interventions, both traditional and digital, have been shown to have similar effects on outcomes as their disorder-specific counterparts [30], yet are more flexible and scalable [31].

MYLO is an artificial intelligence-based conversational agent that emulates the method of levels (MOL) therapy [32], a universal therapeutic approach based on perceptual control theory (PCT), which is a unified model of psychological functioning [33-36]. According to PCT, psychological distress is caused by conflicting goals or values within an individual, and these internal conflicts lead the individual to experience loss of control, which manifests as psychological distress [33]. People have a hierarchy of different goals (values, ideals, and internal standards), with more important goals higher in the hierarchy and unresolved conflicts at higher levels entailing more chronic distress. According to PCT, an in-built learning process called reorganization can resolve conflict when a person’s awareness is sustained on the superordinate goal that drives the conflict. Therefore, therapeutic interventions based on PCT aim to sustain a client’s awareness of their problem to explore the conflict until a superordinate goal enters awareness, which is in turn explored to support the effective reorganization and restoration of control [32,37,38].

MOL therapists encourage clients to freely express and explore their problems by asking questions with appreciative curiosity to sustain a client’s attention to their problems and bring the client’s awareness to background thoughts that emerge while they are talking [32]. MYLO emulates MOL by asking users to describe their problem (eg, “I’m worrying about my daughter’s illness”), by identifying key terms and phrases in the users’ text (eg, “worrying”), and by selecting and generating an appropriate question based on these terms (eg, “What goes through your mind when you worry about this?”). By doing this, MYLO aims to provide a real-time personalized experience to users to help them explore their problems. Therefore, MYLO can address some of the challenges and recommendations previously mentioned regarding conversational agents.

An initial proof-of-concept randomized controlled trial (RCT) compared a single session of a MYLO prototype with a session with ELIZA, a chatbot that uses natural language processing to emulate a human-centered psychotherapist [39] with a university student sample [40]. MYLO was rated as more helpful than ELIZA, and participants in the MYLO group indicated significantly higher rates of problem resolution than those in the ELIZA group (P<.05). A similar, larger RCT with students and staff of 2 universities in the United Kingdom also found that MYLO was rated by users as more helpful than ELIZA [41]. Both studies found that participants reported reduced problem-related distress and reduced symptoms of depression, anxiety, and stress after using both chatbots. However, given that these studies used a single, approximately 20-minute session for university students and staff, clinically significant changes were not expected. A secondary aim of the study by Gaffney et al [40] was to test whether the mechanisms of change and reorganization of conflict described by PCT mediated participants’ helpfulness ratings and clinical outcomes. Indications of the mechanism were coded from the text conversations and were associated with greater distress reduction, improved problem resolution, and more positive expectations of using MYLO.

For the next stage of development, a MYLO prototype was provided for 2 weeks to a community sample of adults with self-reported diagnoses of anxiety or depression [31]. Participants identified the properties of MYLO that they found helpful, including providing a greater sense of control, a sense of being understood and respected, and being a good fit for the individual. The most helpful questions were those that allowed the user to talk freely and gain a new perspective or awareness of their problem.

Although participants have generally found MYLO to be an acceptable intervention, MYLO faces similar challenges to other chatbots, namely, ensuring that the content is appropriate and not repetitive [31]. To address these challenges and improve MYLO, participants from earlier studies made several suggestions for improving the MYLO interface, including modernizing it, using a more traditional messaging app layout, providing crisis contact information, and increasing the diversity and number of questions.

This Study

In response to these recommendations, this study developed a new MYLO progressive web application (PWA) and interface. We recruited a youth advisory committee to help co-design this interface so that it would be accessible, engaging, and appropriate for young people aged 16 to 24 years experiencing symptoms of anxiety, depression, or low mood. To test the feasibility and acceptability of the new interface, we used a protocol similar to that of Gaffney et al [31] and gave participants the new MYLO app to test for 2 weeks, followed by qualitative interviews and focus groups. The results of this study will inform a second developmental stage that will include upgrading MYLO’s database and a fully powered RCT within this population. The specific aims are as follows:

1. Assess the feasibility of recruiting diverse young participants for a research study on MYLO
2. Assess MYLO’s acceptability and gain feedback on the research design
3. Assess the feasibility and acceptability of providing MYLO via a PWA to smartphone users aged 16 to 24 years
4. Assess the preliminary effects of MYLO on target outcomes for a future fully powered trial (eg, problem distress, anxiety, and depression symptoms) and the proposed mechanisms of change (eg, expressing oneself openly and freely and other tendencies toward the reorganization of goal conflict).

Methods

The MYLO Co-Design Phase

At the start of this research project, MYLO was available only as a web application. We recruited a youth advisory committee of 10 young people who had experienced anxiety or depression. A total of 7 committee members attended meetings or provided written feedback during the co-design phase. This group
included 4 nonbinary people, 2 women, and 1 man, aged 16 to 24 years. Of these, 6 members lived in the Perth Metro area and 1 lived in a regional (ie, country) area of Western Australia. The panel was recruited by the lived experience researcher on the team through their existing networks and through the Consumer and Community Involvement program at the first author’s institute.

A total of 4 youth advisory committee meetings were held between July 1, 2022, and October 14, 2022, during which time the youth advisory committee tested different iterations of MYLO and provided feedback that was then presented to the software development team (Multimedia Appendix 1). The software development team implemented the committee’s feedback, and new iterations were then returned to the committee for further feedback.

**Ethics Approval**

Approval for the case series was obtained from the Curtin University Human Research Ethics Committee (HREC2022-0466).

**Recruitment**

A web-based digital advertisement was created and used to advertise the study between September 14, 2022, and October 21, 2022. The advertisement was shared by all members of the research group through their existing networks and personal social media pages. Twitter and Facebook profiles were also created for the MYLO app to advertise the study. The Twitter post shared by the MYLO Twitter profile was retweeted 22 times and gained 1731 impressions, and 15 clicks were gained on the survey link. A targeted Facebook advertising campaign was purchased for a 7-day period between October 12, 2022, and October 19, 2022, with a target audience limited to those in Western Australia aged 16 to 19 years, to recruit more participants aged 20 years. During this time, the advertisement reached 6275 people, resulting in 174 clicks on the survey link. During the recruitment period, several local and state-wide organizations, including consumer advocacy groups, mental health services, and other youth agencies, shared the advertisement either on social media or through their networks.

**Participants**

Inclusion criteria were participants aged 16 to 24 years, currently living in Western Australia, having lived experience of anxiety or depression, having a smartphone and access to the internet, and being able to confidently read and type in English. Participants were also asked if they were able to commit to completing the web-based assessments each week (no more than 30 min/week) and were able to attend the 1-hour focus group after the testing phase. Participants were excluded if they were currently experiencing severe depressive symptoms or frequent suicidal thoughts. This was assessed using the Patient Health Questionnaire-9 [42], and participants who scored 20 (the established threshold for severe depressive symptoms) or scored 2 or 3 on the suicidal thoughts item (item 9) were excluded. All participants aged 18 years were asked if they wanted to provide their parents’ or guardians’ consent, and 3 of the 6 did.

We had several demographic targets to ensure that a wide range of young people were able to test and provide feedback on MYLO. These targets were a minimum of 2 men, 2 women, 2 people who identified as nonbinary, two 16- to 17-year-olds, two 18- to 21-year-olds, two 22- to 24-year-olds, 2 people who identified with a minority cultural group in Australia, and 2 people who lived in rural or remote regions of Western Australia (ie, not within the Perth or Peel metropolitan region). According to the Australian Bureau of Statistics [43], a minority cultural group in Australia is any group other than Australian, any of the North-West European groups, or any of the Southern European Groups (not including South Eastern and Eastern Europeans).

Participants who followed the link or QR code on the advertisement were taken to an expression of interest survey hosted by Qualtrics (Qualtrics International Inc). The survey contained questions to ensure that participants met the inclusion and exclusion criteria, understood the study protocol, and provided informed consent and their contact details. Figure 1 shows the number of participants excluded or lost throughout this process. The research team reviewed the demographic information of the 27 eligible participants who completed the expression of interest survey and contacted a diverse range of young people. In total, 19 people were contacted to participate in the study; of these, 17 completed the baseline survey. A total of 4 participants were identified as completing the baseline survey from outside Australia, and their data were discarded, leaving a final sample of 13 participants.
Materials

Web-Based Survey

Web-based assessments were administered via an anonymous survey hosted by Qualtrics at baseline, after 1 week of testing MYLO (during-testing survey), and after 2 weeks of testing MYLO (posttesting survey). Participants were sent an email or text containing the link to each survey as well as email or text reminders to complete the survey the following day. To link participants’ responses across the 3 time points while retaining anonymity, participants generated a subject-generated identification code [44]. Table 1 provides a summary of the self-report questionnaires included in the web-based assessments. Although we did not expect to see a significant change in these outcomes after 2 weeks of using MYLO, we calculated whether any participants experienced a reliable change in their scores over the 2 weeks. This was calculated using Cronbach $\alpha$ for each questionnaire and the reliable change method described by Evans et al [45]. To assess the acceptability of the questionnaire, participants were asked to rate how easy they thought each self-report questionnaire was on a 5-point scale, ranging from $-2$ (very difficult) to 2 (very easy), and participants could also provide qualitative feedback for each questionnaire via an open text box.

Figure 1. The number of participants excluded or lost through the expression of interest survey. PHQ-9: Patient Health Questionnaire-9.
### Table 1. The questionnaires used in the case series.

<table>
<thead>
<tr>
<th>Questionnaire</th>
<th>Measures</th>
<th>Scoring</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient Health Questionnaire-9 [42]</td>
<td>9 items; depression</td>
<td>0-4: minimal depression, 5-9: mild depression, 10-14: moderate depression, 15-19: moderately severe depression, and 20-27: severe depression.</td>
</tr>
<tr>
<td>General Health Questionnaire-12 [47]</td>
<td>12 items; psychiatric impairment</td>
<td>Traditional (acute) scoring method used. Scores range from 0 to 12, and higher scores indicate a greater possibility of psychological distress.</td>
</tr>
<tr>
<td>Short Form-6D version 2 [48]</td>
<td>6 items; general health</td>
<td>Scores range from ~0.685 to 1, with 1 indicating perfect health. Australian weights were used for this sample.</td>
</tr>
<tr>
<td>Psychological Outcome Profiles [49]</td>
<td>4 items used for scoring; change in problem-related distress over the course of therapy</td>
<td>Each item is scored from 0 (I do not believe this at all) to 100 (I believe this completely). The mean of the 10 items is used as the outcome.</td>
</tr>
<tr>
<td>Reorganization of Conflict Scale [50]</td>
<td>10-item subscale; goal conflict awareness and the proposed mechanism of change in the method of levels therapy</td>
<td>Scores range from 10 to 40. Higher scores indicate higher perceived general self-efficacy.</td>
</tr>
<tr>
<td>General Self-Efficacy Scale [51]</td>
<td>10 items; self-efficacy</td>
<td>Each item is scored from 1 (not at all) to 5 (very much). We calculated the mean scores for the unwanted thoughts, relationship impacts, hindering impacts, understanding, and problem-solving subscales. Item 17 measures “other impacts,” an optional item that is not used in scoring.</td>
</tr>
<tr>
<td>Session Impact Scale [52]</td>
<td>17 items; session (therapeutic) satisfaction</td>
<td>Outcome is a percentile ranking from 0 to 100, with scores &gt;68 considered above average.</td>
</tr>
<tr>
<td>System Usability Scale [53]</td>
<td>10 items; user experience of digital systems</td>
<td>Users select the days they used MYLO (Monday to Sunday), estimated how long an average conversation lasted (in mins), and how many conversations they had on each day they used MYLO.</td>
</tr>
<tr>
<td>User Engagement Survey [a]</td>
<td>3 items; how often and for how long</td>
<td>Users select the days they used MYLO (Monday to Sunday), estimated how long an average conversation lasted (in mins), and how many conversations they had on each day they used MYLO.</td>
</tr>
</tbody>
</table>

*aDenotes surveys that were only presented at the during-testing survey and post-testing survey.

### Manage Your Life Online

MYLO was provided as a PWA that could be accessed through a web browser and downloaded onto the user’s smartphone (Figure 2). From the home page, users could choose to start a new conversation, resume their last conversation, or access a range of mental health resources. When a new conversation commences, MYLO asks the user “Please tell me what’s on your mind.”, users are then able to type free text about the problem they would like to explore. MYLO analyzes users’ text for key terms (eg, “anxious”) and phrases (eg, “can’t sleep”) and responds with a question (eg, “What do you think about feeling anxious?”). These questions are designed to emulate the questions that an MOL therapist would use [32] and aim to prompt users to consider their problems from a higher level of awareness. By doing so, users can become better at resolving their problems and, therefore, reduce the level of problem-related distress they experience [32]. The conversation continues with MYLO asking questions and the user responding until the user chooses to end the conversation.

Within the interface, users also have access to a list of mental health resources as well as a button that connects them to the Lifeline call center—an Australian suicide prevention hotline. These resources were included to provide users with the ability to connect to face-to-face or crisis services if they feel they need to. Users also have limited ways to customize their profile by changing their profile name and the colors of their avatar (their initial on a colored square). Both features were recommended by the youth advisory panel to improve safety and acceptability, respectively. MYLO uses built-in control systems to identify relevant terms in users’ responses and to generate an appropriate question in response, and it uses these systems to improve at both tasks. Users rate each of MYLO’s responses, which generates an error term for each unique term and question pairing as well as each question and term on its own. Each question, term, and question and term pairing started with an error term of 0, meaning they are “helpful” at the beginning of the testing phase. The more a question and term are rated as unhelpful or neither helpful nor unhelpful, the larger their error terms become (with higher error values being added for unhelpful ratings compared with neither helpful nor unhelpful ratings). Equally, the more questions and terms are rated as helpful, their error terms are reduced. Once a question and term pairing has been used >5 times, MYLO uses the error terms to sort its list of possible questions when selecting the best questions, making it less likely that unhelpful questions will be selected and more likely that helpful questions will be selected. It was decided the pairing needed to be used 5 times before learning begins to ensure that error terms were based on a pattern of helpfulness, as a question may be helpful to one person but unhelpful to others. This information was used to examine the engagement of participants with MYLO, explore the acceptability of MYLO’s questioning, and judge whether MYLO can learn and thereby adjust its questioning in the future based on the ratings given by the participants.

https://humanfactors.jmir.org/2023/1/e46849
Focus Group
The topic guide (Multimedia Appendix 2 [31,54]) was adapted from the study by Gaffney et al [31] to investigate what the participants found helpful and unhelpful regarding MYLO. Other questions were adapted from the study by Ly et al [54] to gauge the engagement of participants and use of MYLO. Participants were also asked about their experience of completing the web-based assessments to examine the acceptability of the measures used for future studies. The focus group was recorded and transcribed using an independent local transcription service. Inductive content analysis of the transcripts was conducted by the first author according to the steps described by Years and Gillam [55]. The coding schemas were discussed with the last author and refined.

Procedure
All participants were provided with the newly developed MYLO PWA to test for 2 weeks. During this time, participants completed 3 web-based assessments: at baseline, after 1 week of testing MYLO (during-testing survey), and after 2 weeks of testing MYLO (posttesting survey). The assessments contained several self-report questionnaires on psychological well-being and experience with the MYLO app. After the 2-week testing phase, participants attended a web-based focus group to provide qualitative feedback on their experience with the MYLO PWA and the study protocol. Participants received digital gift vouchers of Aus $20 (US $13.40) per hour (maximum of 4 hours) for their time testing MYLO, completing the web-based assessments, and attending the focus group.

Results
Recruitment and Retention
The final sample consisted of 13 participants who completed the baseline survey. The final sample met all the diversity targets for gender, age, cultural group, and region (refer to the Methods section for more details). The demographics of the participants are summarized in Table 2. Of the 13 participants, 10 (77%) completed all web-based surveys, and the 10 participants provided qualitative feedback (n=5, 38% participants attended a web-based focus group and owing to limited availability, n=3, 23% attended web-based interviews, n=1, 8% provided written feedback to the focus group questions, and n=1, 8% provided brief feedback via email). A total of 15% (2/13) of participants dropped out in the first week of testing (ie, they did not complete the during-testing survey), and neither of these participants gave a reason. The participants who did not complete the final survey informed the researchers that they were too busy; this was also the same participant who provided brief feedback via email. Another participant who provided written feedback rather than attending an interview informed the researchers that they were unwell while the focus groups and interviews were being conducted and therefore could not attend the focus groups and interviews.
We retrospectively collected information on the sample’s sexuality to further assess the diversity of the sample, and of those who disclosed their sexuality, 3 described themselves as heterosexual, 1 as lesbian, 1 as pansexual, and 1 as “vincian/gay (attracted to men and masc. [gender] people).” Participants completed an anonymous survey during the second week of testing MYLO and were asked to self-describe their sexuality. In the future, this information will be gathered during the expression of interest survey.

Acceptability of the Research Design

Web-Based Survey

The difficulty ratings for all the surveys are provided in Multimedia Appendix 3. None of the questionnaires received a negative mean score. The lowest ratings were for the Session Impact Scale (SIS), measuring therapeutic satisfaction (mean 0.3, SD 1.06), and the engagement questionnaire (mean 0.4, SD 1.17), indicating that both were of “neutral” difficulty to complete. The engagement questionnaire was also the only questionnaire to be rated very difficult by 1 participant who explained in the textbox that they had severe memory problems because of a psychological disorder and, therefore, could not remember when they used MYLO during the week. All other questionnaires received mean scores of 0.6 to 0.8, indicating that participants found them neutral to complete. Some participants who completed the web-based difficulty ratings (10/13, 77%) also provided qualitative feedback in the survey (4/10, 40%), with 1 stating that they preferred the Patient Health Questionnaire-9 (depression) style questions to the Short Form-6D version 2 (SF-6Dv2) General Health Questionnaire style questions, although they rated both as very easy. Another participant suggested it would be useful to ask about life events that are impacting the participants to better understand why their scores may have changed during the testing period:

Possibly a useful thing is asking about the context? i.e., Has anything changed in the past few weeks that we should take into consideration when we are evaluating this survey?

Finally, 1 participant used the textbox to state that their health had deteriorated during the testing phase but that it was not MYLO’s fault:

MYLO not helping was not MYLO fault. Bad health and stuff get worse. MYLO did not make it worse.

Participants were also asked to rate the overall survey length. Of the 10 participants who completed the posttesting survey, 7 said the survey was too long and 3 said it was about right (no one said it was too short).

Qualitative Feedback on the Intervention

In focus groups or interviews, participants were positive about their experience of participating in the MYLO study, and some indicated they would be interested in participating again. Length of the surveys and testing time frames were both found to be acceptable. Most participants (7/8, 88%) found the surveys easy to complete, and none of the questions or surveys were flagged as distressing, although some participants (3/8, 38%) described the surveys as “samey” or repetitive. Despite the time commitment, several participants indicated that they saw the value of participation and were happy to contribute. Recommendations and technical issues were also reported to the team and are detailed in Textbox 1.
Textbox 1. Participant recommendations to improve the research design.

- Change the testing time frame so that surveys are completed every 2 weeks.
- Surveys should take a maximum of 15 minutes.
- Conduct short qualitative interviews midway through the testing phase.
- Monitor life events during the testing phase.
- Should be able to pause and resume completing the survey over several sittings.
- One scale (reorganization of conflict) required the participants’ phone to be in landscape mode.
- The slider on the Reorganization of Conflict Scale (0-100) should be changed to a Likert-type scale like the other surveys.

Feasibility and Acceptability of MYLO

We assessed the feasibility and acceptability of MYLO across 3 categories: engagement with MYLO, acceptability of the interface, and acceptability of MYLO’s therapeutic conversations.

Engagement With MYLO

Participants reported using MYLO between 1 and 4 days a week in the first week and having 1 to 3 conversations with MYLO on those days. Participants reported using MYLO for a variety of reasons: when they needed to share or talk about something, when they felt low, and when they had spare time. Several participants attributed their drop in use in the second week to MYLO’s repetitive questioning. Another participant said they forgot about MYLO, and this contributed to their lower use:

Because it didn’t become, like, part of my routine that I do all the time, it just… I’d forget that it was a thing.
[Nonbinary, 16-17 years]

The average length of a conversation ranged from 2 to 30 minutes, with most conversations lasting 10 to 15 minutes (n=8). Of the participants who provided conversation length for each day (n=9), the total time of using MYLO over the week ranged from 7 to 62 minutes, with most participants using MYLO for 30 to 35 minutes (n=6). In the second week, participants reported using MYLO between 1 and 7 days and having 1 to 5 conversations with MYLO on those days. The average length of a conversation ranged from 5 to 15 minutes. Of the participants who provided conversation length for each day (n=7), the total time spent using MYLO over the week ranged from 15 to 40 minutes.

MYLO Use Data

Only conversations with 1 response were included in this analysis. The participants had 32 conversations with MYLO between October 17, 2022, and November 4, 2022. This time is longer than 2 weeks as recruitment of participants was staggered; the final participant finished the 2-week testing phase on November 7, 2022. The word count of these conversations ranged from 58 to 2104 words and participants sent 2 to 20 texts. A total of 13 conversations had 5 participant texts, 11 had between 6 and 10 participant texts, and 8 had between 11 and 20 participant texts. Participants used MYLO at various times of the day: 8 of them used it between midnight and 6 AM, 12 between 6 AM and noon, 12 between noon and 6 PM, and no one used MYLO between 6 PM and midnight. The texts sent by participants included 23 different themes (this does not include the themes from the 13 conversations that were 5 responses, as MYLO does not currently record this information; this also only includes themes that were used to choose a question; other themes may also have been present in texts sent by participants) drawn from 48 unique terms (refer to Multimedia Appendix 4 for the full list).

Participants rated 15 conversations: 6 were rated as helpful, 2 as neither helpful nor unhelpful, and 7 as unhelpful. As shown in Figure 3, in total 100% of the conversations 1000 words were rated as helpful, and all the remaining conversations 1000 words were rated as either unhelpful or neither.

At the end of the testing phase, 61 unique questions and term pairings were rated 75 times by the participants, including 40 unique questions. Table 3 provides a summary of the questions used more than once and the ratings they received during the testing phase. Of the 61 question and term pairings, 41 (67%) had an error term of 0 at the end of the testing phase, indicating that the pairings (and the questions and terms in the pairings) were only ever rated as helpful. The remaining pairings had various error terms 0, indicating that they received ratings other than helpful. The differences in error terms indicate that MYLO records user ratings of the questions and uses this feedback to adjust its learning system.
Table 3. Questions used more than once by Manage Your Life Online and their helpfulness ratings (n=75)\textsuperscript{a}.

<table>
<thead>
<tr>
<th>Questions</th>
<th>Total, n (%)\textsuperscript{b}</th>
<th>Helpful, n (%)\textsuperscript{c}</th>
<th>Neither, n (%)\textsuperscript{c}</th>
<th>Unhelpful, n (%)\textsuperscript{c}</th>
</tr>
</thead>
<tbody>
<tr>
<td>“When you feel &quot;d&quot;, what goes on in your body or in your mind?&quot;</td>
<td>9 (12)</td>
<td>6 (67)</td>
<td>1 (11)</td>
<td>2 (2)</td>
</tr>
<tr>
<td>“What makes you use the word &quot;d&quot;?&quot;</td>
<td>7 (9)</td>
<td>5 (71)</td>
<td>1 (14)</td>
<td>1 (14)</td>
</tr>
<tr>
<td>“How do you think you could begin to do that?&quot;</td>
<td>4 (5)</td>
<td>4 (100)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>“Where do you think this feeling is coming from?&quot;</td>
<td>4 (5)</td>
<td>4 (100)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>“When you say &quot;d&quot;, how does that actually feel for you?&quot;</td>
<td>4 (5)</td>
<td>3 (75)</td>
<td>0 (0)</td>
<td>1 (25)</td>
</tr>
<tr>
<td>“You are saying that you want to do something. What is getting in the way?”</td>
<td>3 (4)</td>
<td>2 (67)</td>
<td>0 (0)</td>
<td>1 (33)</td>
</tr>
<tr>
<td>“When you say &quot;d&quot;, how often do you feel like this?&quot;</td>
<td>3 (4)</td>
<td>3 (100)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>“How is this feeling affecting you?”</td>
<td>3 (4)</td>
<td>3 (100)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>“Tell me more about what you would like?&quot;</td>
<td>3 (4)</td>
<td>2 (67)</td>
<td>1 (33)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>“How do you feel about looking at the future like that?&quot;</td>
<td>3 (4)</td>
<td>2 (67)</td>
<td>1 (33)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>“What would help you achieve that?”</td>
<td>2 (3)</td>
<td>2 (100)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>“What thoughts about yourself are associated with &quot;d&quot;?&quot;</td>
<td>2 (3)</td>
<td>2 (100)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Total number of questions rated by participants during the testing phase.
\textsuperscript{b}Percentage of the total number of questions rated by participants during the testing phase.
\textsuperscript{c}Percentage of times the question was rated by participants during the testing phase.
\textsuperscript{d}Text in quotation indicates the term that was identified in the users’ text.

Acceptability of the Interface

Participants rated the usability of the MYLO interface during testing and posttesting using the System Usability Scale. The mean rating across both time points was 73.57 (SD 16.02) and ranged from 37.50 to 97.50 (median 77.5). Scores increased slightly between the during-testing survey (mean 71.59, SD 16.17) to the posttesting survey (mean 75.75, SD 16.42). The overall mean indicates that MYLO is better than the average of other systems (median score 68 [56]) and is currently ranked within the 65th to 69th percentile of user systems, giving it a grade of B—meaning it is “acceptable” [56]. Single-question scores were examined to determine which areas could be targeted for improvement. Participants only scored 2 questions below average—questions 1 and 5: “I think that I would like to use this system frequently” and “I found the various functions in this system were well integrated,” respectively.

Participants were able to provide brief qualitative feedback after completing the System Usability Scale as well as during the focus groups and interviews regarding the MYLO interface. The participants said that the interface’s simple design made it easy to use. Participants also liked the colors and that MYLO was being developed locally. The participants made
recommendations for MYLO, which are reported under MYLO recommendations in the Results section. Several technical issues were reported but none seemed to cause participants to disengage from using MYLO.

Acceptability of MYLO’s Therapeutic Conversations

Overview

The acceptability of MYLO’s therapeutic conversations with the user was assessed using the participants’ therapist satisfaction scores and focus group and interview data. First, we report participants’ satisfaction with MYLO’s text-based conversations, followed by their satisfaction with the conversations as a therapy session and which aspects of the conversations they felt were useful. Finally, we report the difficulties reported by the participants and their recommendations to improve MYLO.

Satisfaction With MYLO’s Text-Based Conversations

Most participants expressed satisfaction with their conversations with MYLO and liked the text-based conversation system, explaining that it allowed them to access support discreetly without being judged and in different situations. Some participants (2/8, 25%) expressed preferring texting to talking about their feelings:

I prefer, like, texting and getting my feelings out. Just because I can really quickly, like, my fingers catch up to my brain. So, I just prefer the typing. [Woman, 18-21 years]

One interview participant had issues with verbal expression and memory that had previously negatively impacted in-person therapy:

I have some speech issues. So, like, being able to type is a lot easier for me. And it was really good to be able to, like, because that’s an issue that I’ve had with regular therapy as well, like, being able to verbally express. So being able to type everything out was really helpful. So, it was really good in terms of the typing. [Verbal expression issues, nonbinary, 16-17 years]

It would be easy to, like, read back, like, see what I’d said, see what MYLO said, because sometimes, like, in the middle of conversations, I just forget everything, so I have to, like, refresh myself, where was I? And so, it’s really good for that. Like, if I’m in the middle of something, and, you know, we need to go back and get more context, I can. So that’s really helpful, because, again, it was quite a barrier when I was doing in-person therapy where, like, I’d suddenly forget everything in the middle of the session, and I’d have to be, like, “Can you tell me again what we were talking about?” [Memory issues, nonbinary, 16-17 years]

Satisfaction With Therapy Sessions

The mean therapy satisfaction scores across participants were compared with the existing cohorts of participants (Table 4) with anxiety and depression receiving computerized therapy [57] and brief in-person psychological interventions [58,59]. It is worth noting that the participants completed these measures during the weekly surveys rather than after every session with MYLO.

Table 4. Session impact subscale scores for Manage Your Life Online (MYLO) and other psychological therapies.

<table>
<thead>
<tr>
<th>Session impact subscale</th>
<th>MYLO (mean (SD; 95% CI))</th>
<th>Computerized cognitive behavioral therapy [60] (mean (SD; 95% CI))</th>
<th>Therapist-delivered cognitive behavioral therapy [60] (mean (SD; 95% CI))</th>
<th>Psychotherapy [58] (mean (SD; 95% CI))</th>
<th>Psychodynamic therapy [57] (mean (SD; 95% CI))</th>
<th>Cognitive behavioral therapy [57] (mean (SD; 95% CI))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Understanding, mean (SD; 95% CI)</td>
<td>2.43 (1.00; 2.00-2.86)</td>
<td>2.35 (0.49; 1.92-2.78)</td>
<td>3.03 (0.82; 2.23-3.83)</td>
<td>2.60 (1.05; 2.55-2.65)</td>
<td>2.87 (0.71; 2.64-3.10)</td>
<td>2.73 (0.77; 2.48-2.98)</td>
</tr>
<tr>
<td>Problem-solving, mean (SD; 95% CI)</td>
<td>2.14 (1.05; 1.69-2.59)</td>
<td>2.79 (0.76; 2.12-3.46)</td>
<td>3.44 (1.00; 2.46-4.42)</td>
<td>2.87 (1.11; 2.82-2.92)</td>
<td>2.79 (0.64; 2.59-2.99)</td>
<td>3.36 (0.67; 3.14-3.58)</td>
</tr>
<tr>
<td>Relationship, mean (SD; 95% CI)</td>
<td>2.28 (0.93; 1.88-2.68)</td>
<td>2.62 (0.64; 2.06-3.18)</td>
<td>3.43 (0.89; 2.56-4.30)</td>
<td>3.11 (1.04; 3.06-3.16)</td>
<td>3.22 (0.74; 2.99-3.46)</td>
<td>3.28 (0.75; 3.04-3.53)</td>
</tr>
<tr>
<td>Hindering, mean (SD; 95% CI)</td>
<td>1.77 (0.58; 1.52-2.02)</td>
<td>1.19 (0.10; 1.10-1.28)</td>
<td>1.14 (0.15; 0.99-1.23)</td>
<td>1.17 (0.37; 1.15-1.19)</td>
<td>1.20 (0.26; 1.12-1.28)</td>
<td>1.14 (0.28; 1.05-1.23)</td>
</tr>
<tr>
<td>Unwanted thoughts, mean (SD; 95% CI)</td>
<td>1.52 (0.75; 1.20-1.84)</td>
<td>1.35 (0.30; 1.09-1.61)</td>
<td>1.46 (0.32; 1.15-1.77)</td>
<td>1.50 (0.83; 1.46-1.54)</td>
<td>1.51 (0.44; 1.37-1.65)</td>
<td>1.47 (0.49; 1.31-1.63)</td>
</tr>
</tbody>
</table>

Comparison with previous studies suggests that MYLO scored slightly lower on understanding (except when compared with computerized cognitive behavioral therapy [57]), problem-solving, and relationship than the other interventions and slightly higher on hindering impacts. Participants experienced similar unwanted thoughts after using MYLO as after using other interventions. The individual scores across the 5 subscales varied (Multimedia Appendix 5). For example, individual means for understanding and problem-solving (2 subscales that most closely align with MYLO’s proposed mechanisms of change) ranged from 1 to 4. For understanding, 7 participants had a mean of 2, indicating that their sessions with MYLO were at least slightly helpful in supporting them to gain understanding. Similarly, during the focus groups and interviews, many participants described MYLO as helpful and suggested that they were able to gain some insight into themselves or their problems while using it.
The participants said that MYLO made them consider and explore their problems by asking novel questions. Even participants who acknowledged that this was usually a difficult task for them described the process of exploration with MYLO as helpful:

I think it really helped me capture or, like, kind of explore how I felt because usually, what happens is for me, when a problem comes, all my emotions are wrapped in a bundle and it’s hard for me to unravel that, or express or process that. So I think that was helpful. [Woman, 22-24 years]

Some participants (2/8, 25%) stated that since using MYLO they have continued to think about their problems following MYLO’s principles, even when the conversation with MYLO might have ended poorly:

But I’ve noticed, even when I’m not using MYLO, it kind of helps ground me when I’m, like, oh, I need to think about why I’m feeling this way. So, all in all, it has helped me, even today. [Woman, 16-17 years]

I found even when I left the conversation feeling, like, kind of annoyed, I noticed that I would still keep picking up things that MYLO has, like, taught me, especially with kind of trying to understand why I’m feeling something or exactly what I’m feeling, and kind of bringing myself back down to the ground. [Woman, 16-17 years]

Although the comparison of SIS scores suggested that MYLO performed slightly worse than in-person therapy, for some participants (3/8, 38%), the lack of a human therapist improved their experience, as they did not feel judged:

Yeah, and especially because it’s an app, like, I don’t feel judged by anyway. Like, I know, it’s anonymous. [Woman, 18-21 years]

A participant felt that the process was less overwhelming:

Yeah, I think sometimes the presence of someone, like, across from you is, like, overstimulating sometimes, so everything’s, like, going on at once. [Woman, 16-17 years]

**Difficulties With MYLO Conversations**

The SIS scores suggest that all participants experienced some difficulties with MYLO’s conversations, impacting its helpfulness and, in some cases, causing frustration. Some participants (3/8, 38%) felt that MYLO had difficulties understanding them because of how they were typing (ie, number of words and content of the message). Participants who experienced this problem adjusted the language they used, and the problem was resolved:

I had a little bit of an issue at first where I asked, like, I said something specific and it didn’t understand, but once I was using it more, I understood, like, to use broader words, stuff like that. [Woman, 18-21 years]

Some participants (4/8, 50%) also found it difficult to explain their feelings:

I found it difficult when it would ask to kind of, like, explain, like, in a few more sentences what you were feeling because I’m not much of a talker. [Woman, 16-17 years]

Another barrier for participants initially engaging with MYLO and having a successful conversation was their internal state. Participants who were distressed did not want to or did not have the capacity to explore their problems.

The largest problem that caused participants to disengage with MYLO during a conversation was the repetition of questions or the use of very similar questions that made participants feel they were repeating themselves:

That’s why I walked away frustrated, just because it said the same things, and then I didn’t want to have to re-explain myself. Like, I don’t want to expand on what I said because I’ve already just said it. [Woman, 18-21 years]

Questioning was also described by 1 participant as overwhelming. Finally, some participants (4/8, 50%) also had trouble understanding some of MYLO’s questions, so they struggled to answer them:

I don’t know, I sort of struggled with, like, the questions that MYLO asked though, like, I sort of struggled to understand most of them, like, what they were sort of wanting me to talk about, I guess. [Nonbinary, 16-17 years]

When asked to elaborate, the participant described the questions as vague and gave an example:

It would ask me, like, why I said the word “stressed” about something...which I didn’t really know how to answer. [Nonbinary, 16-17 years]

These questions are typical of MOL therapy, where a therapist will inquire about the language or words people use to encourage them to explore their experience without the therapist assuming they understand the client’s experience based on the language a client used [61]. This may be challenging for some users, especially if they have not attended an MOL session before. These types of questions also seem to be those rated “unhelpful” most often (Table 3) and therefore will need to be improved in future development stages.

**MYLO Recommendations**

Participants provided recommendations to address some of these issues and improve other aspects of user experience (Table 5). The suggestions included changes to the MYLO interface that provided more control to the user over the aesthetics of the interface, such as options to customize the colors, changes to the MYLO database (eg, a larger range of questions), and additional features (eg, mindfulness or grounding techniques) to help participants get into the right headspace to use MYLO by reducing their initial distress.
To fulfill the recommendation to save old conversations, participants need to be able to create a unique user profile and log in. Case-series participants were, therefore, asked about different methods of achieving this and their preferences. Participants did not reach a consensus on how best to achieve this, but it was important to all of them that logging in and accessing MYLO remained easy and straightforward. Many participants expressed concerns about remembering passwords or other log-in credentials, especially if they were in an emotional state when they wanted to talk to MYLO. Some participants (2/8, 25%) were concerned about data privacy and indicated that they would want to be advised on how and what data were being stored.

Table 5. Participants’ recommendations to improve Manage Your Life Online (MYLO).

<table>
<thead>
<tr>
<th>Recommendations</th>
<th>Participants (n=10), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants sometimes forgot to talk to MYLO; hence, they would like notifications to use MYLO that they could control the frequency of.</td>
<td>6 (60)</td>
</tr>
<tr>
<td>Participants wanted to be able to revisit the previous conversations and would like to save old conversations, or sections of conversations.</td>
<td>7 (70)</td>
</tr>
<tr>
<td>Participants wanted more control over the look of the app and a way to make it feel like their own space, such as customizable color schemes.</td>
<td>4 (40)</td>
</tr>
<tr>
<td>MYLO has a “Resume/Pause conversation” button, but some participants experienced issues with this system and would like it to be improved.</td>
<td>4 (40)</td>
</tr>
<tr>
<td>Participants wanted a native app that was easier to download and access through their smartphones.</td>
<td>3 (30)</td>
</tr>
<tr>
<td>Participants wanted the option of using speech-to-text to improve their ability to express their feelings or problems.</td>
<td>3 (30)</td>
</tr>
<tr>
<td>Participants wanted the ability to text multiple times in a row rather than having MYLO respond after each message to suit their natural texting behaviors more closely.</td>
<td>3 (30)</td>
</tr>
<tr>
<td>Participants wanted the app to include mindfulness and grounding techniques that they could use if they were too distressed to talk with MYLO.</td>
<td>3 (30)</td>
</tr>
<tr>
<td>Participants wanted some strategies to be recommended for the recurrent problems they discussed with MYLO.</td>
<td>2 (20)</td>
</tr>
<tr>
<td>Participants suggested having rotating backgrounds similar to Windows to improve the aesthetics of MYLO.</td>
<td>2 (20)</td>
</tr>
<tr>
<td>Participants wanted an instructional demonstration of how to use and talk with MYLO to improve its usability.</td>
<td>2 (20)</td>
</tr>
<tr>
<td>One participant wanted the ability to use MYLO offline, improving MYLOs usability and accessibility.</td>
<td>1 (10)</td>
</tr>
<tr>
<td>One participant suggested MYLO be able to use and recognize emojis to communicate with young people more naturally.</td>
<td>1 (10)</td>
</tr>
<tr>
<td>One participant suggested a space in the app to record or vent without MYLO asking questions.</td>
<td>1 (10)</td>
</tr>
<tr>
<td>One participant suggested a cross platform profile so they could use MYLO on any device and access their previous or paused conversations.</td>
<td>1 (10)</td>
</tr>
<tr>
<td>One participant suggested the ability for MYLO to connect users with a person or expert in the app to receive human support.</td>
<td>1 (10)</td>
</tr>
<tr>
<td>One participant suggested a larger repertoire of questions to reduce repetition.</td>
<td>1 (10)</td>
</tr>
<tr>
<td>One participant requested access to peer support within the app.</td>
<td>1 (10)</td>
</tr>
<tr>
<td>One participant wanted examples of how to answer questions in the conversation window.</td>
<td>1 (10)</td>
</tr>
<tr>
<td>One participant suggested MYLO be able to check on users’ well-being during the conversation to ensure they are safe to continue.</td>
<td>1 (10)</td>
</tr>
<tr>
<td>One participant said the “Helpful/neither/unhelpful” buttons needed to be clearer, both what their purpose is and their visibility on the screen.</td>
<td>1 (10)</td>
</tr>
<tr>
<td>One participant suggested that MYLO should be able to give positive feedback during conversations when a user is doing well or progressing.</td>
<td>1 (10)</td>
</tr>
</tbody>
</table>

Participants in each focus group and interview were also asked whether they would recommend MYLO. Of the 8 participants, 6 (75%) would recommend MYLO and the remaining 2 (25%) would recommend MYLO with some improvements. The reasons for participants to recommend MYLO were as follows: MYLO is easy to use (n=2), MYLO is easy to access (n=2), traditional psychological support is expensive (so MYLO would ideally be free, n=2), and MYLO is a good supplementary tool (n=1). One participant said that they knew friends who liked to work through their problems in a similar way:

*I'll definitely be recommending it to my friends and stuff. Because a lot of them process issues the same way I do where you sort of need to, like, talk it out and figure things out for yourself. So, it'd be really helpful for them as well.* [Nonbinary, 16-17 years]
One participant thought it would be particularly useful for young people, and another participant said it would be useful for those who do not feel comfortable accessing in-person therapy and are experiencing milder symptoms.

**Target Outcomes**

As the current sample was small and the testing time frame was short, we did not expect to observe significant improvements in the participants’ clinical outcomes. Table 6 presents the mean scores over time. Cohen $d$ was calculated for each outcome at 2 weeks relative to baseline and showed at least a small effect (ie, Cohen $d$ $\geq 0.2$) for each domain, except for general health, depression, and self-efficacy. The sample’s problem-related distress scores were further examined by calculating the Psychological Outcome Profiles effect size, which provides an estimate of the effect size of change for the sample between baseline and posttesting (ie, pretherapy and posttherapy). The Psychological Outcome Profiles scores; this is calculated by subtracting the mean posttherapy score (posttesting) from the mean pretherapy score (baseline) and dividing the result by the SD of the pretherapy score. In this sample, the effect size was 1.50, indicating a large effect size [59].

Table 7 shows individual changes in scores from the baseline survey to the posttesting survey for participants who completed the measures at both time points. The reliable change index was calculated for each participant on each outcome, and those that were found to have reliably changed are denoted in Table 7. A total of 3 participants reliably deteriorated on a single measure during the testing phase: 1 participant’s general health (SF-6Dv2), 1 participant’s anxiety (Generalized Anxiety Disorder Assessment-7), and 1 participant’s self-efficacy (General Self-Efficacy Scale). Inspection of the participant’s SF-6Dv2 results showed a 1-point deterioration in their scores for physical functioning, body pain, vitality, and mental health between the baseline and posttesting surveys. A total of 7 participants experienced reliable improvements during the testing survey, and at least 1 participant improved in each outcome, with 1 participant improving across all outcomes.

### Table 6. Mean scores on clinical outcomes at baseline, during, and after testing Manage Your Life Online for 2 weeks.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Cronbach $\alpha$</th>
<th>Reliable change index [45]</th>
<th>Participant*</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>General health</td>
<td>.76</td>
<td>0.32</td>
<td>-0.09</td>
<td>0.24</td>
<td>-0.15</td>
<td>0.28</td>
<td>-0.29</td>
<td>0.27</td>
<td>0.34</td>
<td>-0.52</td>
<td>-0.25</td>
<td>-0.03</td>
<td></td>
</tr>
<tr>
<td>Depression</td>
<td>.76</td>
<td>4.23</td>
<td>-4.00</td>
<td>2.00</td>
<td>-1.00</td>
<td>-2.00</td>
<td>4.00</td>
<td>-5.00</td>
<td>3.00</td>
<td>3.00</td>
<td>4.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anxiety</td>
<td>.84</td>
<td>5.15</td>
<td>-1.00</td>
<td>-3.00</td>
<td>2.00</td>
<td>-1.00</td>
<td>2.00</td>
<td>-3.00</td>
<td>-11.00</td>
<td>2.00</td>
<td>5.00</td>
<td>-6.00</td>
<td></td>
</tr>
<tr>
<td>Psychiatric impairment</td>
<td>.84</td>
<td>3.10</td>
<td>-3.00</td>
<td>1.00</td>
<td>2.00</td>
<td>-4.00</td>
<td>2.00</td>
<td>3.00</td>
<td>-9.00</td>
<td>2.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Goal conflict reorganization</td>
<td>.89</td>
<td>13.23</td>
<td>12.70</td>
<td>9.10</td>
<td>-1.00</td>
<td>-6.60</td>
<td>14.20</td>
<td>12.10</td>
<td>33.10</td>
<td>11.40</td>
<td>9.90</td>
<td>-6.10</td>
<td></td>
</tr>
<tr>
<td>Self-efficacy</td>
<td>.81</td>
<td>3.27</td>
<td>5.00b</td>
<td>-3.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>-1.00</td>
<td>8.00</td>
<td>-1.00</td>
<td>-4.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Problem-related distress</td>
<td>.66</td>
<td>3.98</td>
<td>-6.00b</td>
<td>-2.00</td>
<td>-1.00</td>
<td>-3.00</td>
<td>2.00</td>
<td>-8.00</td>
<td>-10.00</td>
<td>-3.00</td>
<td>-8.00</td>
<td>2.00</td>
<td></td>
</tr>
</tbody>
</table>

*The change column presents mean change between baseline and posttesting survey scores; therefore, the scores of the participants who did not complete the posttest survey were not included.

### Table 7. Change from the baseline survey score to the posttesting survey score.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Cronbach $\alpha$</th>
<th>Reliable change index [45]</th>
<th>Participant*</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>General health</td>
<td>.76</td>
<td>0.32</td>
<td>-0.09</td>
<td>0.24</td>
<td>-0.15</td>
<td>0.28</td>
<td>-0.29</td>
<td>0.27</td>
<td>0.34</td>
<td>-0.52</td>
<td>-0.25</td>
<td>-0.03</td>
<td></td>
</tr>
<tr>
<td>Depression</td>
<td>.76</td>
<td>4.23</td>
<td>-4.00</td>
<td>2.00</td>
<td>-1.00</td>
<td>-2.00</td>
<td>4.00</td>
<td>-5.00</td>
<td>3.00</td>
<td>3.00</td>
<td>4.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anxiety</td>
<td>.84</td>
<td>5.15</td>
<td>-1.00</td>
<td>-3.00</td>
<td>2.00</td>
<td>-1.00</td>
<td>2.00</td>
<td>-3.00</td>
<td>-11.00</td>
<td>2.00</td>
<td>5.00</td>
<td>-6.00</td>
<td></td>
</tr>
<tr>
<td>Psychiatric impairment</td>
<td>.84</td>
<td>3.10</td>
<td>-3.00</td>
<td>1.00</td>
<td>2.00</td>
<td>-4.00</td>
<td>2.00</td>
<td>3.00</td>
<td>-9.00</td>
<td>2.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Goal conflict reorganization</td>
<td>.89</td>
<td>13.23</td>
<td>12.70</td>
<td>9.10</td>
<td>-1.00</td>
<td>-6.60</td>
<td>14.20</td>
<td>12.10</td>
<td>33.10</td>
<td>11.40</td>
<td>9.90</td>
<td>-6.10</td>
<td></td>
</tr>
<tr>
<td>Self-efficacy</td>
<td>.81</td>
<td>3.27</td>
<td>5.00b</td>
<td>-3.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>-1.00</td>
<td>8.00</td>
<td>-1.00</td>
<td>-4.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Problem-related distress</td>
<td>.66</td>
<td>3.98</td>
<td>-6.00b</td>
<td>-2.00</td>
<td>-1.00</td>
<td>-3.00</td>
<td>2.00</td>
<td>-8.00</td>
<td>-10.00</td>
<td>-3.00</td>
<td>-8.00</td>
<td>2.00</td>
<td></td>
</tr>
</tbody>
</table>

*The values in the cells under each participant are the changes in their scores on each outcome measure from baseline to posttesting.

bDenotes reliable changes.
Discussion

Principal Findings

In this study, we developed a PWA version of MYLO through iterative co-design with a diverse group of young people from Western Australia. We established the feasibility of our research design to test MYLO by recruiting the target number of participants and reaching our diversity criteria with respect to gender, age, ethnicity, and geographical region. We successfully retained 77% (10/13) of the participants for the web-based surveys and qualitative feedback after 2 weeks. The retention rate, albeit for a short period, compares favorably with similar studies on mental health chatbots [21,24,62]. A good level of retention was consistent with the acceptability of the research design, with most measures rated as easy to complete. We established a good level of acceptability of the app in terms of use, ratings of system utility, therapeutic process, and helpfulness of MYLO’s questions, as well as gathering qualitative data and recommendations to improve MYLO in the future.

By analyzing MYLO’s use of search and question terms, we established that MYLO worked as it had been designed and could learn from users. We also obtained several interface design recommendations to implement in the next developmental stage. The effect sizes for the research measures over the 2-week period varied, but they showed sufficient promise to embark on a larger trial of effectiveness, with a longer intervention period and comparison condition.

By undertaking an iterative co-design phase, we were able to incorporate many user-led features and ensure that the interface was underpinned by expert insight. We have documented recommendations for further development of MYLO, which can also inform mental health chatbots more broadly. For example, users’ request for additional personalization and customization options aligns with previous user experience research, which found that young people prefer apps that they can be personalized and tailored to their needs [63,64]. Despite the growing number of digital mental health interventions and chatbots available for young people, few researchers are engaging with users to improve the effectiveness, uptake, and adherence rates of their innovations [65]. Of the 30 digital mental health technologies identified by Jones et al [65], only 2 papers reported on the co-design of a mental health chatbot [66,67]. By engaging in meaningful co-design and ensuring that MYLO’s interface is engaging and appealing to young people, we believe that we will be able to achieve high levels of retention and engagement with the app, which we would not have otherwise achieved, leading to improved clinical outcomes for users. We plan to test this hypothesis in a larger, fully powered trial.

Achieving a diverse sample is critical for assessing the acceptability of MYLO and the research design. Our findings suggest that MYLO is acceptable for a diverse range of young people. This builds on previous research, as studies involving real-world samples often provide very little information about participants [68]. Other studies have mostly included student populations [19,24,69]. White people [24,70], and women [19,70].

Although most of the web-based measures were rated as easy to complete, the engagement measure and the SIS were rated as “neutral.” Given the apparent issues with participants retrospectively reporting their use of MYLO, these data will be collected directly from the app in the future. To ensure time efficiency, secure user identification was not implemented in this case series. However, this has now been established and will allow a range of anonymized user-specific metrics to be collected from MYLO and analyzed. With regard to the SIS, our findings suggest that MYLO is similar to other digitized mental health interventions [60] but currently performs slightly worse than face-to-face therapy [57,58,60]. Because very few studies have measured therapy or user satisfaction with mental health chatbots, there is no broadly accepted measure for this group of technologies. Using the SIS allowed us to examine the subscales of concepts relevant to MYLO (such as problem-solving) as well as identify where MYLO was performing well and where it could be improved. The session evaluation questionnaire [71] is also used to evaluate face-to-face therapy, but the subscales may be less applicable to chatbots or other digitized therapies that are user led. For example, the smoothness subscale may not fairly evaluate a chatbot or any therapy that does not follow an organized plan but is rather completely user led.

Another key improvement recommended by participants was the addition of a brief measure of adverse life events that occurred during the trial, as some participants experienced stressful life events during the trial that they felt may have impacted their clinical outcome scores. Many trials have already gathered information on adverse life events that occurred before the trial using a variety of measures [72,73]. A recent review [74] only identified 2 controlled trials of chatbots that gathered information on adverse events during the trial but specifically asked about harms caused by the chatbots [25,75]. To our knowledge, no studies have collected data on stressful life events that are not attributed to the intervention being investigated. The planned effectiveness trial will gather the usual safety information, that is, adverse events caused by MYLO, and allow participants to report other events in their life that may have impacted their clinical outcomes. This will allow us to see whether any confounding factors contributed to the results and also to see how MYLO is able to support people while they are experiencing different conditions and levels of stress.

The mixed methods approach to exploring the experience of using MYLO allowed us to gain a well-rounded and in-depth view. For example, we characterized the lengths and timing of conversations, discovered reasons for both its use and lack of use, and established MYLO’s successful use of search terms and related questions and its capacity to learn to adapt these weightings through user ratings. We identified a potential threshold of 1000 words for a conversation with MYLO to be rated as helpful, as opposed to unhelpful or “neither.” Similarly, researchers found that a higher number of messages exchanged with another artificial intelligence chatbot was associated with more positive feedback [76], and increased engagement led to improvements in symptoms of anxiety and depression [77]. We
will attempt to replicate this finding in the full trial and potentially use machine learning to identify the “signatures” of these “long and helpful” conversations. With secure anonymous user identification, tracking individual users across multiple conversations will further improve our understanding of the trajectory of “helpful” conversations.

Both qualitative analyses and quantitative data in this study provided insight into how MYLO was helpful. Participants’ ratings of the therapeutic process with MYLO were comparable with computerized cognitive behavioral therapy on several subscales, although generally less favorable than benchmarked brief, in-person psychotherapies. Most notably, MYLO seemed to approach in-person therapies in terms of ratings of how well it promotes understanding of a problem, but it scored lower in terms of the quality of the relationship. This is expected because MYLO is not currently programmed to try to foster a relationship with the user; rather, its primary aim is to promote the user’s understanding of a problem in greater depth and detail through curious questions. Consistent with this observation, participants reported that they would continue to ask themselves questions similar to those asked by MYLO after leaving a conversation. Indeed, some participants (3/8, 38%) found the lack of a human therapist to be advantageous. These findings are consistent with the theoretical principles of MYLO (PCT), which implies that everyone differs in what external therapeutic conditions allow the internal process of psychological change to occur [78].

Consistent with the abovementioned perspective, we found a small-to-moderate effect size for improvement in “reorganization of conflict,” the proposed mechanism of change, after the 2-week access to MYLO. The large effect size for reducing scores on the primary outcome (problem-related distress) supports this as the primary outcome measure for the planned effectiveness trial. This finding is consistent with earlier studies on the brief use of MYLO [41]. Similar to earlier brief interventions, we did not expect to find substantial effect sizes for clinical measures, and we did not. However, we recognize that these are only within-group effect sizes that have the usual potential biases (eg, maturation effects or attrition bias), but they do provide preliminary evidence for “promise” of MYLO to merit evaluation in an RCT. The planned effectiveness trial will also initially offer MYLO for 3 months rather than for 2 weeks and will use the version of MYLO that will incorporate many of the recommendations that have been generated from this case series and prioritized systematically. The acceptability and feasibility of collecting these clinical data remotely within this age group have, nonetheless, been established.

**Future Developments**

Before the planned effectiveness trial, we will undertake further developments to address user concerns and recommendations. The largest issue raised by users was that sometimes MYLO’s questioning could become repetitive. This is an issue faced by many chatbots [79], and we believe it can be overcome by using natural language processing, such as Chat Generative Pretraining Transformer (ChatGPT). We plan to explore the use of a natural language processing platform that uses a bias engine specifically trained on mental health topics. Using this technology may improve the ability of MYLO to better understand and identify relevant terms in users’ conversations, thereby improving the helpfulness of questions throughout a conversation. Furthermore, this technology could allow MYLO to phrase questions in a variety of novel ways without requiring a very large database of questions.

To address the issue of some participants (4/8, 50%) not understanding the questions, we are exploring 2 strategies. First, we are planning to add a short introduction at the start of each conversation explaining the purpose of the questions and type of questions users can expect during a MYLO conversation. This is a technique used by some practitioners, A Churchman and N Gluckman (meeting, March 2023), when conducting the MOL therapy with young people to prime clients to be open to exploration. We are also exploring ways for users to prompt MYLO to rephrase a question when needed. This includes offering rephrasing and context to questions in tool tips or using a natural language processing platform to generate new questions with the same aim that might be simpler for users to understand.

To improve MYLO’s scores on the session impact subscales (understanding, problem-solving, relationship, and hindering impacts), we are planning to undertake another co-design and development phase to improve MYLO’s ability to support and understand young people. We anticipate that this will involve increasing the range of common problems faced by young people and the range of language (including slang) that MYLO is able to recognize and respond to [80]. We also hope to expand the range of ways MYLO responds, without changing MYLO’s goal of asking curious questions, to include encouragers [81] to help users feel understood [82].

This study has several limitations. First, the case series used a small sample. Therefore, we did not conduct any inferential statistics on the clinical outcome measures and could not make any substantial comments on MYLO’s effectiveness in improving the mental health of young people. The results of this study should be considered with caution, as it is possible that any effects found could be because of the natural recovery processes rather than an impact of MYLO. We aim to address this limitation in a larger trial. Second, the short 2-week follow-up time, although demonstrated a promising impact on problem-related distress, was unlikely to have an impact on anxiety and depressive symptoms. We will offer MYLO for a longer period and anticipate that prolonged decreases in problem-related distress will lead to improvements in anxiety and depression.

**Conclusions**

In conclusion, we developed and tested the feasibility and acceptability of the newly developed version of MYLO, a mental health chatbot app, through iterative co-design with a diverse group of young people from Western Australia. By engaging in a meaningful co-design, the study was able to achieve high levels of retention and engagement, leading to improved clinical outcomes for users. Participants provided several interface design recommendations to further improve MYLO’s acceptability to be implemented in the next developmental stage, including additional personalization and customization options.
Participants’ improvements in their ability to resolve internal conflicts and problem-related distress provided sufficient promise to embark on a larger trial of effectiveness with a longer intervention period.

Acknowledgments

This study has been supported by the Western Australian Department of Health through the Future Health Research and Innovation Fund. The authors would like to thank Jason Wright, Leah Riungu-Kalliosaari, Dan Marrable, the youth advisory committee, and the study participants for their contributions to this study. Their expertise, feedback, and time were invaluable in making Manage Your Life Online what it is today.

Authors’ Contributions

All authors contributed to study conception and design. ARWH, GA, and JD contributed to data collection. ARWH and WM contributed to the analysis and interpretation of the results. ARWH and WM contributed to manuscript preparation. All authors reviewed the results and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Co-design iterative process.
[PDF File (Adobe PDF File), 876 KB - humanfactors_v10i1e46849_app1.pdf]

Multimedia Appendix 2
Focus group topic guide.
[DOCX File, 21 KB - humanfactors_v10i1e46849_app2.docx]

Multimedia Appendix 3
Difficulty ratings.
[PDF File (Adobe PDF File), 265 KB - humanfactors_v10i1e46849_app3.pdf]

Multimedia Appendix 4
Summary of themes and terms used in the testing phase.
[DOCX File, 22 KB - humanfactors_v10i1e46849_app4.docx]

Multimedia Appendix 5
Individual Session Impact Scale subscale scores.
[DOCX File, 28 KB - humanfactors_v10i1e46849_app5.docx]

References

5. 9 in 10 Aussie teens now have a mobile (and most are already on to their second or subsequent handset). Roy Morgan. 2016 Aug 22. URL: https://www.roymorgan.com/findings/9-in-10-australian-teenagers-now-have-a-mobile-and-most-are-already-on-to-their-second-or-subsequent-handset [accessed 2023-05-26]


Abbreviations

- MOL: method of levels
- MYLO: Manage Your Life Online
- PCT: perceptual control theory
- PWA: progressive web application
- RCT: randomized controlled trial
- SF-6Dv2: Short Form-6D version 2
- SIS: Session Impact Scale
Correction: Performance of a Web-Based Reference Database With Natural Language Searching Capabilities: Usability Evaluation of DynaMed and Micromedex With Watson

Angela Rui1*, MA; Pamela M Garabedian2*, MS; Marlika Marceau2, BA; Ania Syrowatka1,3, PhD; Lynn A Volk2, MHS; Heba H Edrees1,3,4, PharmD; Diane L Seger2, RPh; Mary G Amato1,4, BCPS, MPH, PharmD; Jacob Cambre1, BS; Sevan Dulgarian1, BA, BS; Lisa P Newmark2, BA; Karen C Nanji2,3,5, MPH, MD; Petra Schultz6, BS, PharmD; Gretchen Purcell Jackson7,6, MD, PhD; Ronen Rozenblum1,3, MPH, PhD; David W Bates1,2,3,9, MS, MD

1Division of General Internal Medicine, Brigham and Women's Hospital, Boston, MA, United States
2Clinical and Quality Analysis, Mass General Brigham, Somerville, MA, United States
3Harvard Medical School, Boston, MA, United States
4Massachusetts College of Pharmacy and Health Sciences (MCPHS), Boston, MA, United States
5Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, Boston, MA, United States
6Merative, Cambridge, MA, United States
7Vanderbilt University Medical Center, Nashville, TN, United States
8Intuitive Surgical, Sunnyvale, CA, United States
9Harvard TH Chan School of Public Health, Boston, MA, United States
*the authors contributed equally

Corresponding Author:
Angela Rui, MA
Division of General Internal Medicine
Brigham and Women's Hospital
75 Francis St
Boston, MA, 02115
United States
Phone: 1 978 397 0082
Email: arui@partners.org

Related Article:
Correction of: https://humanfactors.jmir.org/2023/1/e48468

(JMIR Hum Factors 2023;10:e48468) doi:10.2196/48468

In “Performance of a Web-Based Reference Database With Natural Language Searching Capabilities: Usability Evaluation of DynaMed and Micromedex With Watson” (JMIR Res Protoc 2023;25:e47678) the authors noted two errors.

1. The Authors Contributions section currently reads as:

All authors contributed to the study conception; design; and acquisition, analysis, or interpretation of the data. AR, PMG, AS, LAV, DLS, GPJ, and DWB were responsible for study conception or design. PMG, HHE, DLS, and MGA developed the interview guides. MM, AS, SD, and LPN conducted participant recruitment. PMG acted as the interview moderator and had either AR or MM assisting with data collection during testing. PMG, MM, JC, and SD abstracted the data from interview recordings. Data analysis was performed by PMG, MM, and AR. The first draft of the manuscript was written by AR and PMG, with all authors reviewing the draft and providing critical feedback. All authors contributed to and approved the final manuscript.

And will be changed to:

RR and DWB are co-senior authors and contributed equally.

All authors contributed to the study conception; design; and acquisition, analysis, or interpretation of the data. AR, PMG, AS, LAV, DLS, GPJ, and DWB were responsible for study conception or design. PMG, HHE, DLS, and MGA developed the interview guides. MM, AS, SD, and LPN conducted participant recruitment. PMG acted as the interview moderator and had either AR or MM assisting with data collection during testing. PMG, MM, JC, and SD
abstracted the data from interview recordings. Data analysis was performed by PMG, MM, and AR. The first draft of the manuscript was written by AR and PMG, with all authors reviewing the draft and providing critical feedback. All authors contributed to and approved the final manuscript.

2. In the original article, the ORCID number for Petra Schultz was reported as follows:

0000-0001-7337-1046

And has been updated to:

0000-0001-7949-9243

The correction will appear in the online version of the paper on the JMIR Publications website on May 18, 2023, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 24.04.23; this is a non-peer-reviewed article; accepted 02.05.23; published 18.05.23.

Please cite as:
Correction: Performance of a Web-Based Reference Database With Natural Language Searching Capabilities: Usability Evaluation of DynaMed and Micromedex With Watson
JMIR Hum Factors 2023;10:e48468
URL: https://humanfactors.jmir.org/2023/1/e48468
doi:10.2196/48468
PMID:37201180

©Angela Rui, Pamela M Garabedian, Marlika Marceau, Ania Syrowatka, Lynn A Volk, Heba H Edrees, Diane L Seger, Mary G Amato, Jacob Cambre, Sevan Dulgarian, Lisa P Newmark, Karen C Nanji, Petra Schultz, Gretchen Purcell Jackson, Ronen Rozenblum, David W Bates. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 18.05.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Selection of and Response to Physical Activity–Based Social Comparisons in a Digital Environment: Series of Daily Assessment Studies

Abstract

Background: Innovative approaches are needed to understand barriers to and facilitators of physical activity among insufficiently active adults. Although social comparison processes (ie, self-evaluations relative to others) are often used to motivate physical activity in digital environments, user preferences and responses to comparison information are poorly understood.

Objective: We used an iterative approach to better understand users’ selection of comparison targets, how they interacted with their selected targets, and how they responded to these targets.

Methods: Across 3 studies, different samples of insufficiently active college students used the Fitbit system (Fitbit LLC) to track their steps per day as well as a separate, adaptive web platform each day for 7 to 9 days (N=112). The adaptive platform was designed with different layouts for each study; each allowed participants to select their preferred comparison target from various sets of options, view the desired amount of information about their selected target, and rate their physical activity motivation before and after viewing information about their selected target. Targets were presented as achieving physical activity at various levels below and above their own, which were accessed via the Fitbit system each day. We examined the types of comparison target selections, time spent viewing and number of elements viewed for each type of target, and day-level associations between comparison selections and physical activity outcomes (motivation and behavior).

Results: Study 1 (n=5) demonstrated that the new web platform could be used as intended and that participants’ interactions with the platform (ie, the type of target selected, the time spent viewing the selected target’s profile, and the number of profile elements viewed) varied across the days. Studies 2 (n=53) and 3 (n=54) replicated these findings; in both studies, age was positively associated with time spent viewing the selected target’s profile and the number of profile elements viewed. Across all studies, upward targets (who had more steps per day than the participant) were selected more often than downward targets (who had fewer steps per day than the participant), although only a subset of either type of target selection was associated with benefits for physical activity motivation or behavior.

Conclusions: Capturing physical activity–based social comparison preferences is feasible in an adaptive digital environment, and day-to-day differences in preferences for social comparison targets are associated with day-to-day changes in physical activity motivation and behavior. Findings show that participants only sometimes focus on the comparison opportunities that support their physical activity motivation or behavior, which helps explain previous, equivocal findings regarding the benefits of physical activity–based comparisons. Additional investigation of day-level determinants of comparison selections and responses is needed to fully understand how best to harness comparison processes in digital tools to promote physical activity.
Introduction

Background
Engaging in regular physical activity (PA) has wide-ranging and meaningful benefits for physical and mental health [1-3]. Although activity of moderate to vigorous intensity confers unique cardiovascular protection [4], lighter-intensity activity is linked to positive outcomes and is recommended to promote health [5,6]. Conversely, physical inactivity is a key contributor to many of the leading causes of death in the United States and worldwide, including cardiovascular disease and cancer [7-9].

Identifying determinants of PA engagement has been a research priority for several decades and has informed a myriad of prevention and intervention efforts [10]. However, despite these efforts, adults in the United States rarely engage in sufficient PA to protect their health; recent estimates indicate that only 50% meet recommended levels of PA [11], although estimates vary by calculation approach [12]. Consequently, there is a clear need for work that can offer additional insights into PA barriers and facilitators—particularly those that could inform PA promotion efforts on a large scale.

Digital tools such as web platforms and mobile apps show promise for maximizing accessibility to PA resources as they are available for use as needed and can respond to varying contexts in daily life. Specifically, these tools can harness the power of the social environment to support PA by connecting individuals with other users without requiring synchronous interaction [13]. For example, social comparison processes can be activated by sharing PA data between users as captured by a wearable monitor [14]. Exposure to others’ PA behavior allows users to evaluate their own PA relative to that of others [15] using features such as leaderboards and competitive challenges [16,17]. Upward comparison, via exposure to someone doing better with PA (eg, with more steps per day), can inspire the comparer to reach the upward target’s level and provide guidance for how to achieve a similar outcome [18]. Downward comparison, via exposure to someone doing worse with PA (eg, with fewer steps per day), can prompt the comparer to avoid becoming like the downward target to maintain their status [19,20]. Social comparison is expected to work in these ways to motivate users to maintain or increase their PA [21,22].

As a result, features of digital PA tools that activate social comparison processes are popular and have received considerable attention [14,23]. Literature in this area shows some evidence that social comparisons affect PA motivation and behavior (via digital tools and more broadly [24-26]). For instance, direct access to information about others’ PA behavior results in attending more group exercise classes than access to discussions with others about PA (to facilitate social support [27]). However, the effects of comparisons in both upward and downward directions on PA outcomes are heterogeneous and poorly understood. Some people experience decreased PA motivation or behavior in response to social comparisons, including those that are self-selected from a range of options [27-29].

Furthermore, responses to comparisons of PA (with respect to motivation and behavior) are not necessarily consistent for the same person across time; a person may respond positively at some times and negatively at others depending on the daily context [30]. In addition to the direction of a comparison (upward vs downward), a feature that may affect a comparison’s proximal influence on PA outcomes is its scale, or the relative distance between the comparer and target. Comparisons to others doing just a little bit better or worse than the self may have the biggest impact as the target’s outcome seems achievable (upward) or imminent (downward) and the comparer is motivated to improve or maintain their status [15,31,32]. In contrast, comparisons to others who are doing much better or worse may be demotivating as the target’s outcome seems unattainable (upward) or unlikely (downward).

Despite the ubiquity of social comparison features in digital tools to promote PA, the optimal approach to activating comparison processes in a digital environment is not clear. Allowing users to select their preferred comparison target appears to be more effective for promoting PA than restricting exposure to a single (nonpreferred) target [33], and many digital comparison opportunities allow the user to select or focus on a subset of targets from a range of options (eg, leaderboards). However, as noted, even self-selection often results in negative responses. Specifically, there is a current need for additional insights into users’ comparison selections, their interactions with these selections, and the extent to which users respond positively (vs negatively) to their selections in a digital PA environment.

Aims of This Study
Given the availability of digital features that activate social comparison processes to promote PA and the equivocal nature of evidence in this area, there is a need for an improved understanding of PA-based comparison selections and responses in a digital environment. Additional information in this domain could elucidate the nature of PA-based comparison processes and help identify the comparisons that are associated with benefits for PA outcomes (vs harms). The aims of this study were to describe PA-based comparison selections (direction and scale) and examine day-level associations between comparison selections and PA outcomes (motivation and behavior), both overall and for within-person differences across days. To achieve these aims, we used data from an existing 3-study series that allowed participants to select a PA-based comparison target from different sets of options with respect to direction and scale. PA motivation was assessed both before and after comparison exposure each day, and PA behavior was captured in steps per day using the Fitbit platform (Fitbit LLC).

social comparison; physical activity; motivation; web platform; selection; exercise; fitness; mobile phone

Introduction
Methods

Study Series Overview
As part of a larger series of studies to investigate the potential for personalizing social comparison opportunities in the context of a social exergame [34-37], participants in each study completed 7 to 28 total days of data collection. In studies 2 and 3, the first 9 days constituted an exploratory period during which all participants selected from various sets of comparison options; the following days introduced a personalized experimental manipulation for half of the participants based on random assignment. This report describes a set of secondary analyses that examine comparison selections, interactions with these selections via a web platform, and associated consequences for PA motivation and behavior during only the initial 7- or 9-day exploratory period in each study.

Recruitment and Eligibility

Consistent Components Across Studies
Across studies, participants were recruited from the Drexel University undergraduate student participant pool using both in-class recruiting and a web-based study scheduling platform (Sona Systems). Students were eligible to participate if they were aged ≥ 18 years, had daily access to a desktop or laptop computer, self-reported that PA was important to them, and had access to a Fitbit account or were willing to create one. Use of either a Fitbit wearable device or the Fitbit smartphone app was acceptable. Students were excluded if they had a medical condition that limited their ability to engage in moderate- or vigorous-intensity PA or were under medical advisement to avoid moderate or vigorous PA.

Participants—Study 1
Of the 11 undergraduate students who expressed interest in participating, 6 (55%) enrolled in this initial pilot phase. In total, 17% (1/6) of the participants did not complete any days of data collection and were excluded, resulting in a sample of 5 students. The average participant took 4690 (SE 1767.99) steps per day during the study period. All participants were undergraduate students aged ≥18 years; however, further demographic data were not collected during this initial pilot.

Participants—Study 2
Through rolling recruitment over the course of 2 months, 119 students expressed interest in participating. Of these 119 students, 66 (55.5%) did not complete the required days of data collection, resulting in 53 (44.5%) participants who enrolled in study 2. The sample comprised 57% (30/53) women and was racially representative of an undergraduate population, with most participants identifying as White (28/53, 53%) or Asian (13/53, 25%; see Table 1 for further demographic information). The average participant took 6376 (SE 351.43) steps per day during the baseline study period.

Participants—Study 3
Through rolling recruitment over 3 months, 90 students expressed interest in participating. Of these 90 students, 35 (39%) did not complete the required days of data collection, resulting in 54 (60%) participants who enrolled in study 3. Most of the participants were women (37/54, 69%) and the majority of students identified as White (23/54, 43%) or Asian (22/54, 41%).

| Table 1. Demographic characteristics of each sample (N=112). |
|-----------------|-----------------|-----------------|
| Gender, n (%)   | Study 1 (n=5)   | Study 2 (n=53)  | Study 3 (n=54) |
| Women           | 30 (57)         | 37 (69)         |
| Men             | 23 (43)         | 17 (31)         |
| Age (years), mean (SD; range) | 22.45 (7.40; 18-53) | 20.31 (2.93;18-36) |
| Race, n (%)     |                 |                 |                 |
| White           | 28 (53)         | 23 (43)         |                 |
| Asian           | 13 (25)         | 22 (41)         |                 |
| Multiracial     | 5 (9)           | 4 (7)           |                 |
| Black           | 4 (8)           | 2 (4)           |                 |
| Other           | 2 (4)           | 2 (4)           |                 |
| American Indian or Alaska Native | 1 (2)       | 0 (0)           |                 |
| Prefer not to say| 0 (0)         | 1 (2)           |                 |
| Ethnicity, n (%)|                 |                 |                 |
| Hispanic or Latino | 3 (6)       | 7 (13)          |                 |
| Not Hispanic or Latino | 49 (92)  | 47 (87)         |                 |
| Not reported    | 1 (2)           | 0 (0)           |                 |

Note: Demographic data were not collected for study 1. Not available.
41%; see Table 1 for further details). The average participant took 3609 (SE 339.32) steps per day during the baseline study period.

**Measures**

**Social Comparison Selections**

As described in the following sections, participants were asked to select user profiles to view each day from a range of options that represented upward and downward comparisons relative to their own PA behavior. They could select multiple profiles each day to view partial information but could only select 1 profile to view in full. Telemetry built into the web application tracked the participants’ navigation of the web app, including the profiles they viewed (in part or in full), the time spent viewing profiles, and the fields they chose to observe for their full selected profile. Comparison selections were operationally defined with respect to the total number each day, the time spent viewing profiles, the number of profile elements viewed, and the direction and scale of the profile selected for full viewing.

**Motivation to Exercise**

Participants in studies 2 and 3 self-reported their immediate motivation to exercise at the start and end of their participation each day (ie, before and after their comparison selections and exposure). Responses to the following statement—“Overall, I would rate my current motivation to exercise as...” —were rated on a scale from 1 (very low motivation) to 5 (very high motivation) at each time point. This approach to assessing motivation was guided by previous work in this area, including prior work by the investigators [28,38].

**PA Behavior**

To maximize accessibility, activity behavior was defined as total steps per day; steps are a commonly used metric to evaluate PA behavior and are associated with health outcomes [39]. Daily step count totals were measured using data pulled from either a Fitbit wearable device or the Fitbit MobileTrack smartphone app. The app is synced to a participant’s accelerometer on their smartphone, which shows validity for assessing steps across devices and operating systems [40]. Of note, we allowed for heterogeneity in the device used to measure daily steps to enhance the generalizability of findings across individuals with and without the means to purchase a wrist-worn device. This approach has been used in prior work, which shows that Fitbit devices and the MobileTrack app do not generate meaningfully different step estimates [41]. Fitbit step data from the previous day were synced with the study website and then displayed to participants when they logged into the study platform each day.

**Procedures**

After completing a web-based screening survey to determine eligibility, eligible individuals provided electronic informed consent and were then directed to a second web page where they completed a battery of global self-report questionnaires (not included in this report). Participants were then given a username and log-in for the daily web-based activity, where on first log-in, they were directed to authenticate a Fitbit account with our web platform so that daily steps could be retrieved. Starting the following day (which allowed for the sign-up day’s steps to be used in the first session), the user was introduced to the relevant activities described in the following sections. Users were asked to log in and complete a session once per day; the time of day was not specified.

Upon log-in, the web server queried the user’s steps for the previous day via the Fitbit application programming interface (API). If it was detected via the API call that Fitbit did not yet have a full account of the previous day’s steps, the web application directed the participant to open the Fitbit app on their mobile device to prompt a data upload. Of note for study 2, there was a short period during data collection (3 days) in which the Fitbit server was not properly syncing with the study website. As a result, participants’ steps displayed upon logging in represented steps from the last successful sync rather than from the previous day’s true step count. This error was remedied on the day it was identified.

**Daily Social Comparison Task**

**Overview**

As in several previous studies, opportunities to make social comparisons came through viewing profiles of individuals described as similar to the participant [42]. After completing the motivation assessment, participants in each study had the opportunity to select one or more profiles to view. These profiles described other individuals who had recently engaged in more or less PA than the participant to represent upward or downward comparison targets at a range of distances from the participant’s own recent PA behavior. Profile options included only minimal information, including only their username (eg, “dmf25”) and step total. Participants were able to click on multiple selections to learn additional information but could only select 1 profile to view in full.

**Study 1**

Study 1 was designed as a proof-of-concept pilot to ensure that the systems worked correctly and that the platform could detect participants’ navigation behavior. Participants were asked to engage in a 5-minute session on the web platform once per day for 7 days. After logging in each day, participants were greeted with their own step total for the previous day, as tracked by their Fitbit device or app. This was posted next to 4 profiles of “other users,” which were created by the system; 2 presented upward comparisons (ie, with step totals of 110% and 130% of the participant’s steps from the day before), and 2 presented downward comparisons (ie, with step totals of 90% and 70% of the participant’s steps from the day before; Figure 1). In each case, a margin of –2% to +2% was applied as noise to protect against potential identification of the study’s aim.

As noted, participants could select multiple profiles to learn additional information about the users, including their city of residence and favorite location to exercise (as shown in Figure 2). However, they would have to select 1 profile to view in full to complete the task for the day. Upon selecting a profile to view in full, participants viewed a page containing a user’s demographics (eg, age, sex, and profession), physical appearance (eg, height and weight), exercise preferences (eg, preferred forms of PA), and other personal information (eg, hobbies; Figure 3).
**Figure 1.** View of the study web page that included 4 comparison targets to select from.

![Figure 1](image1)

**Figure 2.** View of the Overview study web page, in which a profile has been initially selected but not yet selected to view in full. Participants could still go back and peruse other profiles to select from before selecting their final profile for full details (comparison target). Avg: average.

![Figure 2](image2)
Study 2
The goal of study 2 was to examine patterns of user profile selection (ie, comparison targets) and response with respect to PA motivation and behavior. A revised web platform facilitated engagement in a daily, 2-minute task involving the selection of potential social comparison targets (9 days total). After logging in each day, participants viewed a page displaying their step count from the previous day (as collected from the Fitbit API either via a Fitbit wearable device or a smartphone step tracker synced to the Fitbit app). After reporting their initial motivation to exercise (1-5 rating scale), participants were presented with 4 profiles of other “users” of the application, as in study 1. However, instead of offering a consistent set of profiles with respect to step total (ie, 70%, 90%, 110%, and 130% of the participant’s own steps), participants were assigned to one of the following profile sets each day: (1) all 4 profiles lower than the participant’s (downward options only) at 90%, 80%, 70%, and 60% of the participant’s own step total from the previous day; (2) a mix of profiles—2 downward (lower than the participant’s own step total from the previous day at 90% and 80%) and 2 upward (higher than the participant’s own step total from the previous day at 110% and 120%); and (3) all 4 profiles higher than the participant’s (upward options only) at 110%, 120%, 130%, and 140% of the participant’s own step total from the previous day.

In each case, a margin of –2% to +2% was applied as noise to protect against potential identification of the study’s aim. After viewing their selected full profile, participants were asked to report their exercise motivation a second time (1-5 rating scale).

Study 3
The purpose of study 3 was to examine the translation of the profile selection platform to a gamified context, whereby participants were assigned to teams of 3 users. A further revised version of the web application allowed participants to view other users’ PA behavior and personal information (representing comparison targets) using a new format. As in study 2, participants were asked to log in and report their initial exercise motivation (1-5 rating scale). They then viewed brief descriptions of 2 additional profiles (as opposed to 4 in studies 1 and 2) in leaderboard format and were asked to select 1 to view additional information (Figure 4).

After selecting a profile, participants could view a subset of personal information (Figure 5): this view retained their own step total from the previous day to facilitate comparison with the selected user. Participants could access a full Details page once they selected a final profile to view in full.

However, unlike in the previous studies, step totals for other users in study 3 included data from other participants completing their data collection at the same time (ie, user data that were not created by the platform). Each participant was randomly assigned to a team with another user who began the study at the
same time; these participants each saw the other’s step totals as 1 of their 2 profile options. The third user profile displayed in each session was generated and assigned by the platform, selected from the following options: (1) the third profile showed a step total 20% lower than the lower of the 2 live participants, and the individual participant had either the most steps or was in the middle; (2) the third profile showed a step total between that of the 2 live participants, and the individual participant had either the most or the least steps; and (3) the third profile showed a step total 20% higher than the higher of the 2 live participants, and the individual participant had either the least steps or was in the middle.

In each case, a random noise factor of −2% to +2% was added to obscure our process. This approach was designed to test manipulations of the game environment for the 2 live participant teammates by showing a fabricated third user who might provide an optimal comparison experience for the live teammates.

Across the studies, the distances between the user’s steps and the target’s steps (eg, 80% and 140%) were guided by the principle of offering a realistic range of options and by relevant literature. Specifically, there is evidence supporting the Köhler effect and “motivation gain” in a team game environment that shows that participants’ performance improves with a teammate who performs approximately 20% better than they do [43,44]. Under conditions in which users in this study saw both upward and downward targets as options, −20% was offered for symmetry. Other options were selected to retain realism while capturing distances from the user’s own steps that would be perceptible and large enough to show differences in associations with motivation or behavior. In study 3, the design particulars (ie, percentages below, between, or above 2 real users) resulted in a larger range and set of targets. A summary of each study design is presented in Table 2.

**Figure 4.** Options for selecting from 2 user profiles, listing them and the user in descending order and representing their step totals visually (ie, a leaderboard format).

**Figure 5.** Initial profile view in study 3. Avg: average.
scale of comparison targets (profiles) selected (all studies), the randomized set of targets (studies 2 and 3), with comparison target direction and identified at the grand mean) differentially predicted racial/ethnic identification, and age (age treated as continuous in all models. Motivation was not assessed in study 1; total steps per day were assessed in all 3 studies. Additional data were removed from relevant analyses where unreasonable values were observed, including values for time spent viewing profiles (>6 minutes; 4 observations) and steps per day (<100; 38 observations). The resulting data sets for studies 2 and 3 included 472 and 387 observations, respectively. These data sets afforded power of >0.80 for the primary, within-person tests described in this section (α of .05 [45]), although we emphasize effect sizes throughout—PA is described in steps per day, and all other associations are described using semipartial correlation coefficients (sr). Between-person tests were included to describe potential trends only as power was limited by modest sample sizes.

We first used empty models to calculate intraclass correlation coefficients (ICCs) to determine the proportion of variance attributable to-between-person stability in the outcomes of interest. This included participant navigation behavior when interacting with comparison target profiles (time spent viewing the selected profile and number of elements viewed) and PA outcomes (motivation to exercise and steps per day), which were treated as continuous in all models. Motivation was not assessed in study 1; total steps per day were assessed in all 3 studies. Change in motivation in studies 2 and 3 was calculated by subtracting motivation before profile selection from motivation after selection.

Our first aim was to describe PA-based comparison selections, including participant navigation of the platform and the comparison direction and scale of the selected profile. To address this aim, we initially examined whether gender, racial/ethnic identification, and age (age treated as continuous and centered at the grand mean) differentially predicted navigation behavior. We then used descriptives to examine the frequencies of user profile selections in categories, representing the user’s steps as a percentage of the participant’s steps from the previous day (rounded to the nearest 10). The direction and scale of comparison targets (profiles) selected (all studies), the direction or directions of targets presented using randomization (studies 2 and 3), and whether the selected profile represented the other active participant or the fabricated user (study 3) were treated as categorical and subsequently used as predictors of PA outcomes.

Our second aim was to examine day-level associations between comparison selections and PA outcomes (motivation to exercise and steps per day). Analyses used multilevel modeling techniques using SAS PROC MIXED with restricted maximum likelihood estimation to address the nested data structure (ie, days nested within individuals). Gender, racial/ethnic identification, and age were used as covariates in all multilevel models (studies 2 and 3), with comparison target direction and scale (all studies), the randomized set of targets (studies 2 and 3), and fabricated user versus not (study 3) as predictors of PA outcomes. Although users accessed the platform at a range of times across the days of observation in each study, sensitivity analyses showed that the time of day at which users accessed the platform was not associated with any of our outcomes of interest and did not meaningfully change the results or conclusions reported in the next section. For parsimony, we reported the results of all tests without time of day as an additional covariate.

Finally, new navigation behavior and motivation variables were created for studies 2 and 3: between- and within-person variance were distinguished by calculating each person’s mean across days (between-person) and the difference between this person’s mean and the response on a given day (within-person; ie, person-mean centering [46]). This allowed for testing whether steps per day were associated with within-person fluctuation in navigation behavior or motivation, controlling for typical navigation behavior or typical change in motivation from before to after comparison.

**Ethics Approval**
All procedures were approved by the institutional review board of Drexel University (approval 1901006917).

**Informed Consent and Compensation**
All participants provided documentation of informed consent. Compensation for participation was provided through either

### Table 2. Summary of design distinctions between the studies in this series (k=3).

<table>
<thead>
<tr>
<th>Design feature</th>
<th>Study 1</th>
<th>Study 2</th>
<th>Study 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presentation format</td>
<td>Unordered list; user’s own steps from the day before next to list of target options</td>
<td>Unordered list; user’s own steps from the day before next to list of target options</td>
<td>Leaderboard (user ranked in descending order against 2 others)</td>
</tr>
<tr>
<td>Comparison target options presented each day, n</td>
<td>4</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Range of comparison target distance from the user (steps per day), %</td>
<td>70-130</td>
<td>60-140</td>
<td>0-&gt;2000</td>
</tr>
<tr>
<td>Daily condition assignments</td>
<td>None—same set of target options presented each day (2 upward and 2 downward); step totals differed based on user’s own steps per day</td>
<td>Randomized to 1 of 3 sets of targets—upward targets only, downward targets only, or mixed (2 upward and 2 downward); step totals differed based on user’s own steps per day</td>
<td>Randomized to 1 of 3 sets of targets—upward targets only, downward targets only, or mixed (1 upward and 1 downward)</td>
</tr>
</tbody>
</table>

**Statistical Analyses**
All analyses were conducted using SAS (version 9.4; SAS Institute). Missing data were minimal; data were missing for 20% (7/35) of days in study 1 (because of low compliance from 1 participant), 1% (5/477) of days in study 2, and 4.9% (24/486) of days in study 3. Additional data were removed from relevant analyses where unreasonable values were observed, including values for time spent viewing profiles (>6 minutes; 4 observations) and steps per day (<100; 38 observations). The resulting data sets for studies 2 and 3 included 472 and 387 observations, respectively. These data sets afforded power of >0.80 for the primary, within-person tests described in this section (α of .05 [45]), although we emphasize effect sizes throughout—PA is described in steps per day, and all other associations are described using semipartial correlation coefficients (sr). Between-person tests were included to describe potential trends only as power was limited by modest sample sizes.

We first used empty models to calculate intraclass correlation coefficients (ICCs) to determine the proportion of variance attributable to-between-person stability in the outcomes of interest. This included participant navigation behavior when interacting with comparison target profiles (time spent viewing the selected profile and number of elements viewed) and PA outcomes (motivation to exercise and steps per day), which were treated as continuous in all models. Motivation was not assessed in study 1; total steps per day were assessed in all 3 studies. Change in motivation in studies 2 and 3 was calculated by subtracting motivation before profile selection from motivation after selection.

Our first aim was to describe PA-based comparison selections, including participant navigation of the platform and the comparison direction and scale of the selected profile. To address this aim, we initially examined whether gender, racial/ethnic identification, and age (age treated as continuous and centered at the grand mean) differentially predicted navigation behavior. We then used descriptives to examine the frequencies of user profile selections in categories, representing the user’s steps as a percentage of the participant’s steps from the previous day (rounded to the nearest 10). The direction and scale of comparison targets (profiles) selected (all studies), the direction or directions of targets presented using randomization (studies 2 and 3), and whether the selected profile represented the other active participant or the fabricated user (study 3) were treated as categorical and subsequently used as predictors of PA outcomes.

Our second aim was to examine day-level associations between comparison selections and PA outcomes (motivation to exercise and steps per day). Analyses used multilevel modeling techniques using SAS PROC MIXED with restricted maximum likelihood estimation to address the nested data structure (ie, days nested within individuals). Gender, racial/ethnic identification, and age were used as covariates in all multilevel models (studies 2 and 3), with comparison target direction and scale (all studies), the randomized set of targets (studies 2 and 3), and fabricated user versus not (study 3) as predictors of PA outcomes. Although users accessed the platform at a range of times across the days of observation in each study, sensitivity analyses showed that the time of day at which users accessed the platform was not associated with any of our outcomes of interest and did not meaningfully change the results or conclusions reported in the next section. For parsimony, we reported the results of all tests without time of day as an additional covariate.

Finally, new navigation behavior and motivation variables were created for studies 2 and 3: between- and within-person variance were distinguished by calculating each person’s mean across days (between-person) and the difference between this person’s mean and the response on a given day (within-person; ie, person-mean centering [46]). This allowed for testing whether steps per day were associated with within-person fluctuation in navigation behavior or motivation, controlling for typical navigation behavior or typical change in motivation from before to after comparison.

**Ethics Approval**
All procedures were approved by the institutional review board of Drexel University (approval 1901006917).

**Informed Consent and Compensation**
All participants provided documentation of informed consent. Compensation for participation was provided through either
extra credit in college courses or electronic gift cards depending on individual preference.

Results

Study 1

Of the 5 individuals who participated in the initial proof-of-concept test, 4 (80%) completed the expected daily uses of the web platform (ie, 6–7 within 19 days of enrollment); 1 (20%) participant completed 2 daily uses during the allotted time frame. Participants elected to view the full profile for the first user they selected on 71% (20/28 selections) of days. Across days, participants spent an average of 40 (range 3.3–145) seconds on their selected full profile and clicked on an average of 5 (range 0–29) profile elements. Less than 40% of the variability in both the amount of time each participant spent on their selected profiles and the number of elements they elected to view was attributable to stable, between-person differences (ICC=0.28 and 0.36, respectively), suggesting considerable within-person variability in these behaviors across days (P<.001 in all cases).

Table 3. Steps per day by profile (comparison target) selection; percentages represent the step totals of the selected profile relative to the participant’s steps from the previous day rounded to the nearest 10% (n=28).

<table>
<thead>
<tr>
<th>Type of target</th>
<th>Frequency, n (%)</th>
<th>Steps per day, B (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>70%</td>
<td>3 (11)</td>
<td>4023.82 (2927.76)</td>
</tr>
<tr>
<td>90%</td>
<td>4 (14)</td>
<td>1448.51 (2665.40)</td>
</tr>
<tr>
<td>110%</td>
<td>8 (29)</td>
<td>7152.59 (2321.05)</td>
</tr>
<tr>
<td>130%</td>
<td>14 (50)</td>
<td>6081.24 (2050.10)</td>
</tr>
<tr>
<td>Downward (70% or 90%)</td>
<td>7 (25)</td>
<td>2241.73 (2358.08)</td>
</tr>
<tr>
<td>Upward (110% or 130%)</td>
<td>21 (75)</td>
<td>6403.27 (2015.87)</td>
</tr>
</tbody>
</table>

Study 2

Similar to study 1, participants elected to view the full profile for the first user they selected on the vast majority of days (425/472, 90% of selections). Across days, participants spent an average of 18 (range 1.4–130) seconds on their selected full profile and clicked on an average of 9 (range 0–64) profile elements. Most of the variability in both the amount of time each participant spent with their selected profiles and the number of elements they elected to view was attributable to stable, between-person differences (ICC=0.53 and 0.63, respectively), although both showed evidence of fluctuation for the same person across days (P<.001 for both within-person variance components). Men spent slightly longer viewing each profile and selected to view more profile elements than women (P=.09 and P=.13, respectively); both behaviors were also positively associated with age (P=.02 and P=.02, respectively). However, neither time spent viewing nor the number of elements selected meaningfully differed based on racial/ethnic identification, the set of profile options presented, or the type of target selected (P=.63, P=.11, P=.39, P=.36, P=.91, P=.56, respectively).

Upward comparison target selections were slightly more frequent than downward comparison target selections, representing 54.2% (258/476) of all final profile selections. However, overall, the most popular comparison target selection for viewing the full profile were downward targets at 90% of the participant’s steps from the previous day (Table 4). On days when only downward target options were presented, participants most often selected the target with the step count closest to their own (ie, 90% of their steps from the previous day); this trend was reversed on days when only upward target options were presented (ie, 140% of their steps from the previous day, the farthest from their own). When presented with both upward and downward target options, they selected the target with the highest overall step count (ie, 120% of their steps from the previous day).

Average change in motivation from before to after selection was slightly positive across the days (B=0.10, SE 0.05), with considerable within-person variability (ICC=0.18). The lowest increases in motivation occurred on days when only downward target options were presented (Table 4). Interestingly, participants showed decreases in motivation to exercise only on days when they selected targets with 60% and 110% of their own steps from the previous day (Table 4). These represented the farthest downward and closest upward targets from their own steps, respectively. Participants showed increases in motivation on days when they selected all other targets (contrast F=5.38; P=.02; s=0.32), and this trend did not change when controlling for the set of target options shown.
With respect to steps per day, participants took approximately 540 fewer steps on days when both downward and upward target selections were presented relative to only upward or only downward targets (contrast $F_{41}=-3.80; P=.05$). Steps were highest on days when participants selected targets most distant from themselves in both directions—they took approximately 725 more steps on days when they selected targets with 60% and 140% of their own steps from the previous day relative to targets closer to their own steps (contrast $F_{40}=3.76; P=.05$).

As noted, participants did not always select targets that led to increases in motivation to exercise. Within-person, neither motivation nor steps differed based on the amount of time spent viewing the selected profile or the number of profile elements viewed $(P=.28, P=.21, P=.81, P=.90$, respectively). However, controlling for their typical change in motivation to exercise from before to after comparison, on days when participants were more (vs less) motivated than usual after viewing their selected target, they engaged in more steps $(F_{1,418}=9.24; P=.003)$.

### Table 4. Motivation to exercise and steps per day by profile (comparison target) selection in study 2; percentages represent the step totals of the selected profile relative to the participant’s steps from the previous day rounded to the nearest 10% (n=472).

<table>
<thead>
<tr>
<th>Type of target selected</th>
<th>Frequency, n (%)</th>
<th>Change in motivation to exercise, B (SE)</th>
<th>Steps per day, B (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>60%</td>
<td>34 (7.2)</td>
<td>-0.04 (0.11)</td>
<td>6932.36 (597.06)</td>
</tr>
<tr>
<td>70%</td>
<td>32 (7.2)</td>
<td>0.11 (0.11)</td>
<td>6215.89 (605.29)</td>
</tr>
<tr>
<td>80%</td>
<td>51 (10.8)</td>
<td>0.21 (0.09)</td>
<td>5697.40 (515.19)</td>
</tr>
<tr>
<td>90%</td>
<td>100 (21.2)</td>
<td>0.02 (0.07)</td>
<td>6356.75 (418.57)</td>
</tr>
<tr>
<td>110%</td>
<td>84 (17.8)</td>
<td>-0.01 (0.08)</td>
<td>6078.63 (444.95)</td>
</tr>
<tr>
<td>120%</td>
<td>93 (19.7)</td>
<td>0.14 (0.07)</td>
<td>6447.72 (424.89)</td>
</tr>
<tr>
<td>130%</td>
<td>21 (4.4)</td>
<td>0.20 (0.14)</td>
<td>6515.59 (733.75)</td>
</tr>
<tr>
<td>140%</td>
<td>60 (12.7)</td>
<td>0.19 (0.09)</td>
<td>6965.72 (733.75)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type or types of target options shown</th>
<th>Frequency, n (%)</th>
<th>Change in motivation to exercise, B (SE)</th>
<th>Steps per day, B (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Downward only</td>
<td>159 (33.7)</td>
<td>0.04 (0.06)</td>
<td>6573.20 (367.93)</td>
</tr>
<tr>
<td>Downward and upward (2 each)</td>
<td>159 (33.7)</td>
<td>0.12 (0.06)</td>
<td>6020.79 (366.82)</td>
</tr>
<tr>
<td>Upward only</td>
<td>159 (33.7)</td>
<td>0.11 (0.06)</td>
<td>6556.29 (368.49)</td>
</tr>
</tbody>
</table>

### Study 3

Participants elected to view the full profile for the first user they selected on 96.9% (375/387 selections) of occasions. Across days, participants spent an average of 72 (range 1-351) seconds on their selected full profile and clicked to view an average of 12 (range 0-54) profile elements. As in study 2, although the amount of time each participant spent with their selected profiles and the number of elements they elected to view were fairly stable (ICC=0.58 and 0.65, respectively), they showed some variation for the same person across days (within-person variance components; $P<.001$ in all cases). The time spent viewing profiles and the number of profile elements selected were again positively associated with age ($P=.04$ and $P=.03$, respectively), although neither behavior was associated with the set of profile options presented, whether the selected profile represented an upward or downward target, or whether the selected profile was of the fabricated user versus the real participant ($P=.63$, $P=.75$, $P=.88$, $P=.92$, $P=.14$, $P=.80$, respectively). However, unlike in study 2, neither the amount of time spent on the selected profile nor the number of profile elements selected differed by gender or racial/ethnic identification ($P=.93$, $P=.34$, $P=.93$, $P=.35$, respectively).

The method used to generate profiles in study 3 resulted in participant selections of comparison targets ranging from 0% to 20,610% of their steps from the previous day. This represented selections of users with step totals ranging from 0 to 21,132 steps, with 88 selections of users who had <1000 steps and 27 selections of users with >10,000 steps. This generated >90 individual categories of selection, with most of these categories representing upward targets (ie, the selected users had more steps than the participants on the previous day). For ease of interpretation, upward selections were recategorized by percentages of the participants’ steps, as shown in Table 5. Participants selected the fabricated user on most days (210/387, 54.3%); they were more likely to choose the fabricated user when they selected upward (vs downward) targets ($F_{1,330}=4.44; P=.04$) and were least likely to choose the fabricated user when that user was shown as last on the leaderboard ($F_{2,338}=10.20; P<.001$).

As in studies 1 and 2, upward selections were more frequent than downward selections and represented 57.1% (221/387) of all targets selected. However, unlike in study 2, the most popular choice overall was upward at 120% of the participants’ steps from the previous day (55/387, 14.2% of selections; Table 5). Users with 120% of the participants’ steps from the previous day represented 41.4% (53/128) of all selections on days when the fabricated participant was at the top of the leaderboard but <1% (1/127, 0.8% and 1/132, 0.8%) of selections on days when the fabricated user was second or third. Close in overall frequency of selections were users with 80% of the participant’s steps (as in study 2; 41/387, 10.6% of selections) and users with 200% to 999% of the participant’s steps (41/387, 10.6% of
selections). Of note, selecting to view the profile for a user with the same number of steps the participant had on the previous day occurred on 2.3% (9/387) of the days.

Table 5. Change in motivation to exercise before to after profile (comparison target) selection and steps per day by profile selection in study 3; percentages represent the step totals of the selected profile relative to the participant’s steps from the previous day rounded to the nearest 10% (n=387).

<table>
<thead>
<tr>
<th>Type of target</th>
<th>Frequency, n (%)</th>
<th>Change in motivation to exercise, B (SE)</th>
<th>Steps per day, B (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td>10 (2.6)</td>
<td>-0.07 (0.27)</td>
<td>3838.82 (1160.46)</td>
</tr>
<tr>
<td>10%</td>
<td>8 (2.1)</td>
<td>-0.09 (0.30)</td>
<td>2635.07 (1137.18)</td>
</tr>
<tr>
<td>20%</td>
<td>13 (3.4)</td>
<td>0.06 (0.24)</td>
<td>3964.32 (922.67)</td>
</tr>
<tr>
<td>30%</td>
<td>6 (1.6)</td>
<td>0.18 (0.35)</td>
<td>3697.40 (1277.26)</td>
</tr>
<tr>
<td>40%</td>
<td>5 (1.3)</td>
<td>-0.74 (0.38)</td>
<td>3263.22 (1391.01)</td>
</tr>
<tr>
<td>50%</td>
<td>21 (5.4)</td>
<td>0.24 (0.20)</td>
<td>3004.35 (874.38)</td>
</tr>
<tr>
<td>60%</td>
<td>21 (5.4)</td>
<td>0.39 (0.20)</td>
<td>3404.35 (1127.69)</td>
</tr>
<tr>
<td>70%</td>
<td>19 (4.9)</td>
<td>0.50 (0.21)</td>
<td>3221.15 (1130.97)</td>
</tr>
<tr>
<td>80%</td>
<td>41 (10.6)</td>
<td>-0.03 (0.16)</td>
<td>2243.21 (901.90)</td>
</tr>
<tr>
<td>90%</td>
<td>22 (5.7)</td>
<td>0.10 (0.20)</td>
<td>3440.08 (877.97)</td>
</tr>
<tr>
<td>100%</td>
<td>9 (2.3)</td>
<td>0.32 (0.29)</td>
<td>3337.31 (1432.54)</td>
</tr>
<tr>
<td>110%</td>
<td>14 (3.6)</td>
<td>0.07 (0.24)</td>
<td>3598.14 (1053.33)</td>
</tr>
<tr>
<td>120%</td>
<td>55 (14.2)</td>
<td>-0.09 (0.14)</td>
<td>3454.85 (747.45)</td>
</tr>
<tr>
<td>130%</td>
<td>9 (2.3)</td>
<td>0.28 (0.29)</td>
<td>3221.15 (1130.97)</td>
</tr>
<tr>
<td>140%</td>
<td>11 (2.8)</td>
<td>-0.11 (0.26)</td>
<td>2422.84 (1127.69)</td>
</tr>
<tr>
<td>150%</td>
<td>4 (1)</td>
<td>0.16 (0.42)</td>
<td>3448.54 (1551.14)</td>
</tr>
<tr>
<td>160%</td>
<td>6 (1.6)</td>
<td>-0.25 (0.35)</td>
<td>3891.26 (1299.39)</td>
</tr>
<tr>
<td>170%</td>
<td>10 (2.6)</td>
<td>0.06 (0.27)</td>
<td>3345.57 (1093.40)</td>
</tr>
<tr>
<td>180%</td>
<td>5 (1.3)</td>
<td>0.14 (0.38)</td>
<td>5536.70 (1382.12)</td>
</tr>
<tr>
<td>190%</td>
<td>6 (1.6)</td>
<td>0.51 (0.35)</td>
<td>1878.49 (1378.49)</td>
</tr>
<tr>
<td>200%</td>
<td>5 (1.3)</td>
<td>0.04 (0.38)</td>
<td>2900.83 (1372.52)</td>
</tr>
<tr>
<td>110%–199%</td>
<td>18 (4.7)</td>
<td>-0.12 (0.21)</td>
<td>3675.08 (938.78)</td>
</tr>
<tr>
<td>200%–999%</td>
<td>41 (10.6)</td>
<td>0.29 (0.16)</td>
<td>2949.21 (781.46)</td>
</tr>
<tr>
<td>1000%–1999%</td>
<td>8 (2.1)</td>
<td>-0.10 (0.31)</td>
<td>3915.49 (1299.92)</td>
</tr>
<tr>
<td>&gt;2000%</td>
<td>20 (5.2)</td>
<td>-0.16 (0.21)</td>
<td>3771.33 (962.49)</td>
</tr>
</tbody>
</table>

Type or types of target options shown

- Participant either first or second on leaderboard (fabricated user was third or last): 127 (32.8) 0.05 (0.11) 3510.65 (667.93)
- Participant either first or third (last) on leaderboard (fabricated user was second): 132 (34.1) 0.17 (0.11) 3033.56 (669.24)
- Participant either second or third (last) on leaderboard (fabricated user was first): 128 (33.1) 0.02 (0.11) 3573.50 (668.77)

Selected fabricated user

- No: 177 (45.7) 0.06 (0.11) 3248.29 (653.82)
- Yes: 210 (54.3) 0.09 (0.10) 3463.86 (642.90)

Average change in motivation to exercise from before to after selection was again positive across days but extremely small (B=0.08, SE 0.51), although within-person variability was predominant (ICC=0.04). Increases in motivation were largest on days when participants selected users with 190% of their steps from the previous day, followed by users with 70% of their steps from the previous day (Table 5). Participants’ motivation decreased on days when they selected upward targets with steps farthest from their own (ie, >2000% of their steps from the previous day) as well as on days when they selected
users with 10%, 40%, 80%, 120%, and 160% of their steps from the previous day; the greatest decreases were seen on days with selections of 40% of the participants’ own steps from the previous day. Change in motivation was highest on days when the fabricated user was placed between a given participant and the other real participant on the leaderboard relative to days when the fabricated user appeared above or below both real participants (contrast $F_{335}=2.34; P=.12, sr=0.17$). Change in motivation did not meaningfully differ between days when participants selected an upward or downward target (collapsed across percentage categories; $F_{335}=97; P=.34$) or between days when they selected the fabricated user versus the other live participant ($F_{335}=0.00; P=.98$).

With respect to steps per day, participants took approximately 500 fewer steps on days when the fabricated user was placed between themselves and the other real participant on the leaderboard relative to days when the fabricated user appeared above or below both real participants (contrast $F_{303}=2.89; P=.09$). Steps did not meaningfully differ between days when participants did and did not select to view the profile of the fabricated user ($F_{303}=56; P=.46$). Although steps also did not differ overall based on the comparison direction and scale of the selected profile ($P=.90, P=.99$, respectively), interestingly, steps were highest on days when participants selected users with 180% of their own steps from the previous day (approximately 5500 steps) and lowest on days when they selected users with 190% of their own steps from the previous day (approximately 1900 steps; Table 5). Steps also did not meaningfully differ between days when participants selected an upward versus a downward target (collapsed across percentage categories; $P=.90$).

Neither motivation nor steps were associated with daily fluctuation in the amount of time each participant spent on their selected profiles or the number of elements they elected to view (within-person; $P=.60, P=.64, P=.38, P=.34$, respectively). Finally, although the within-person association between participants’ motivation and steps per day was not significant ($F_{335}=1.11; P=.29$), it was noteworthy that the direction of the association was negative—unlike in study 2, on days when they were more motivated than usual after viewing their selected profile, participants took fewer steps than usual ($B=-186.84, SE 177.65$).

**Discussion**

**Principal Findings**

Social comparison processes can be activated to promote PA in digital environments, although individuals’ interactions with and responses to self-selected comparison targets in this context are poorly understood. As social comparison features are already built into many existing digital PA tools [14,16,23], this series of studies was designed to provide additional information about this important aspect of digital PA promotion. We created unique web-based platforms to capture individuals’ selections of social comparison targets, their interactions with information about the selected targets, and their subjective responses to the selected targets over 7 to 9 days, as well as their PA behavior on each of these days. We observed several similarities and differences across these studies that can shed additional light on this area.

First, participants chose to view the full profile of the first participant they selected on the vast majority of days (71%–97%), although many participants explored other profiles before returning to and settling on the first one they had selected. Participants also interacted with the platform and their selected profiles differently across days. They did not merely settle into a pattern of the same behavior each day despite the consistency and simplicity of the task. This underscores the appeal of PA-based comparisons and their potential to sustain engagement with digital tools, although additional testing over longer periods is needed.

Second, in both studies where demographic information was collected, older participants spent more time viewing profiles and selected more profile elements to view than younger participants. This stands in contrast to existing cross-sectional evidence, which suggests that older people are less interested in comparisons than younger people [47]. It is possible that our findings reflect a general tendency among older people to pay more attention to their participation in research than younger people [48]. Alternatively, it is possible that cross-sectional, retrospective self-evaluations of comparison activity do not align with observable behavior; this potential discrepancy is worthy of further investigation given that social comparison is often captured using global self-report measures [49,50]. Also noteworthy is that, although the participants’ ages in these studies ranged from 18 to 56 years, we recruited students enrolled in college who were predominantly in their early 20s. As such, associations with age warrant further investigation. Other observations of differences in behavioral interactions with social comparison information based on demographics (eg, gender) were not consistent across the studies in this series, although the power for these comparisons was limited.

Third, across all studies, the profiles of upward comparison targets were selected for full viewing more often than those of downward comparison targets. This was not an artifact of randomized exposure—each participant had an equal number of opportunities to select upward and downward targets. Moreover, participants tended to select upward targets that were distant from themselves (ie, those who had many more steps than they had) rather than upward targets closer to themselves. Selecting to make upward comparisons, particularly when a range of options is available, is often motivated by a desire for self-improvement [51,52]. Given that participants in these studies indicated that PA is important to them, selecting targets doing extremely well with PA offered an opportunity to learn information from that target that could support achievement of a similar high status [53]. For example, participants could learn new ways to be active from the profiles of very active participants, giving them opportunities to set PA goals to model the target.

However, despite the relative popularity of upward targets, participants also frequently selected downward targets and tended to select downward targets close in steps to their own (vs more distant from their own). Self-selection of downward
targets is often motivated by a desire for self-enhancement [51,52]; seeing oneself as doing better than someone else in a valued domain can be satisfying and provide an emotional boost. The variety of selections across days may indicate day-to-day variability in participants’ needs and immediate goals that could be met with comparison opportunities [54,55].

Importantly, participants did not always select the target that was most useful with respect to either subjective PA motivation or PA behavior—many selections were associated with decreases in motivation, low PA engagement, or both. Similarly, a participant’s change in PA motivation as a result of viewing their selected comparison target was not consistently associated with their PA behavior. Subsets of previous work in this area show important aspects of comparisons that may help explain these findings and, thus, warrant further consideration. One is that people do not always select the comparison opportunities that fulfill either self-improvement or self-enhancement goals; at times, their intentions are to confirm that their own situation is bad or could worsen or to justify not making difficult behavior changes such as increasing their PA (eg, “I’m already doing better than someone else, so I’m doing fine”) [56,57]. Even when they do have positive, goal-oriented intentions for selecting particular comparison opportunities (eg, to learn important information or to feel better), their expectations are not always met by the target provided [58]. In such situations, the comparison opportunity may actually lead to negative outcomes.

In addition, the affective consequences and behavioral correlates of a social comparison selection opportunity may depend on how the comparer interprets the information they receive. The Identification-Contrast Model of comparison processes [59] proposes that the comparer can focus on either similarities or differences between themselves and a target (reflecting identification with vs contrast against the target, respectively). Identifying with an upward target highlights the possibility that the comparer can achieve similar (better) outcomes, and contrasting against a downward target highlights the comparer’s current success (as the outcome could be worse). Conversely, identifying with a downward target suggests that the comparer’s situation is bad or may become worse; contrasting against an upward target highlights the comparer’s inferiority and suggests that the likelihood of achieving similar success is low. In the context of PA and similar comparisons of health behaviors, there is recent evidence showing that greater (vs less) identification with active others is associated with more frequent attendance to exercise classes [60], and identification and contrast processes moderate the association between the type of target selected (upward vs downward) and motivation to engage in healthy behavior [28]. Identification and contrast with respect to both upward and downward comparisons are also known to differ between people and show evidence of fluctuation for the same person over time [61-63]. Thus, in this series of studies, the high day-to-day variability in participants’ PA outcomes that were not fully explained by the direction or scale of the selected target may be due to individual or day-level differences in the extent of identification or contrast with the target. Assessment of these processes in future work could more fully explicate the complexity of social comparison and its optimal use to promote PA engagement. As discussed further in this section, to effectively isolate the source of this variability, removing potential noise coming from variability in the time of day of social comparison selections and exposure would be optimal in future studies.

Finally, we observed differences in findings between studies that may generate additional hypotheses to be tested in future work. For example, PA motivation in response to viewing the selected comparison target was positively associated with within-person behavior in study 2 but not in study 3. Study 2 presented the list of target selection options and the selected target’s step total side by side with the participant’s step total from the previous day. In contrast, study 3 presented social comparison target selection options in a leaderboard format such that the participant saw a visual representation of their rank against the 2 other users. These differences may affect the psychological dynamics of comparison selections and their associations with PA motivation and behavior, in general or for specific individuals. Target selection options in study 3 also included both a real participant and a fabricated user, where the ultimate goal was to determine the optimal placement of the fabricated user to balance the comparison effects on both of the real users. In this study, PA motivation increased the most on days when the fabricated user was in the middle of the leaderboard (between the 2 real users), but steps were lowest on these days. The leaderboard and balance approach may have blunted the potential negative effects of comparisons but also blunted some positive effects.

Participants who enrolled in study 2 were also noticeably more active than those who enrolled in study 3 (and study 1); relative to the US guideline of achieving 10,000 steps per day [6], the average activity level was moderate in study 2 and low in study 3 (and study 1). It is possible that the general correspondence between PA motivation and behavior is stronger for those who are moderately active than for those who are inactive in that those who are moderately active are better able to enact their PA motivation. Distinctions between studies could be due to participant characteristics, study design, or a combination of both. As a result, it is not yet clear whether one study design is more useful than another for activating beneficial PA-based social comparisons or whether there is a subset for whom one is superior to another.

Strengths and Limitations of This Research

This series of studies has several strengths. Specifically, all 3 studies used objective assessment of comparison target (profile) selection, interactions with the target (ie, time spent viewing and number of profile elements viewed), and PA behavior (steps per day) across several days. Studies 2 and 3 also captured motivation to exercise both before and after target selection using a momentary item that was tested in previous work [28,38]. Retention of enrolled participants was high across studies, with minimal missing data. In addition, we used a multilevel analytic approach that allowed for maximizing the utility of intensive repeated assessments, with insights into daily behavior across participants as well as within-person associations across days. Finally, we took an iterative approach such that the platforms used in each study were slightly different
with respect to the comparison target options to allow for preliminary comparisons between and across studies. Although the sample sizes in each study were modest and do not afford definitive conclusions about the sources of divergent results, observations of consistency and inconsistency across studies provide a strong foundation for hypothesis-driven research on a larger scale.

In addition to modest sample sizes, several other limitations are noteworthy. Participants’ access to the web platform was not restricted to a particular time of day or constrained to be consistent for the same participant across days. Consequently, participants may have taken part at varying times of day (eg, before vs midway through vs after engaging in most of their steps for that day). Although participants’ comparisons were anchored to their steps for the previous day, which were already completed, and controlling for time of day did not alter our findings, this inconsistency could mask any effects of social comparison selections on motivation or PA behavior for the current day by allowing for considerable noise between and within participants. In addition, the precision of PA behavior captured likely varied by participant as some used wearable PA monitors (eg, Fitbit wristbands) whereas others used less sensitive smartphone accelerometers. Assessment of PA motivation and behavior was also misaligned—motivation referred to “exercise” (ie, structured bouts of sustained, moderate– to vigorous–intensity movement), and behavior was captured with respect to steps (ie, overall movement at any intensity, including light activity). Although motivation did predict within-person behavior in study 2, this discrepancy may further help explain the lack of association in study 3. Future work should ensure that assessments of cognitive determinants of PA and PA behavior refer to the same behavioral outcomes.

Finally, participants were all students enrolled in college courses who reported that PA was important to them. This ensured that the dimension of comparison (PA) was relevant to the participants [15]. The average participant in each study also fell far short of US recommendations for PA behavior (ie, 10,000 steps per day), suggesting that participants generally represented individuals who could benefit from increasing PA—a target population of interest. However, recruitment from college courses and requiring participants to endorse a preexisting interest in PA resulted in samples of well-educated, motivated, and predominantly White young adults. As noted, there is existing evidence indicating that younger adults report more interest in and show stronger responses to social comparison information than older adults [47]. This may limit the effectiveness of social comparison processes as a PA promotion tool for younger adults, who already tend to be more active than older adults in the United States [11]. These are common problems in digital health research, particularly early-stage work. Additional attention needs to be paid to recruiting and retaining diverse samples to fully understand the range of PA social comparison preferences and responses that may be useful for promoting PA.

Conclusions

Despite these limitations, these findings have several important implications. With respect to platform interface design, users show interest in viewing the profiles of other users and engage with profile content when the initial information available offers social comparison opportunities. Furthermore, as social comparison target selections are often not associated with benefits for PA motivation or behavior, the current real-world conditions for digital PA promotion tools (which offer unrestricted access to other users [14]) do not appear to meet users’ needs. Outcomes could be improved with subtle manipulation of comparison target options. These exploratory findings show that constraining users’ PA-based social comparison options and changing their options across days (with respect to direction and scale) is both feasible and acceptable, with high completion rates. An important next step is to identify the people and immediate contexts for which certain types of comparisons are optimal (eg, older vs younger adults, men vs women, or high vs low precomparison motivation) to allow for systems to offer the PA-based social comparison opportunities that are most likely to benefit users in their daily lives.

Acknowledgments

This work was partially supported by the US National Science Foundation under award 1816470 (principal investigator: JZ) and the Novo Nordisk Foundation Grant under award NNF20OC0066119. Support for the first author’s time was provided by the US National Institutes of Health under award K23HL136657 (principal investigator: DA). The authors would like to thank Thomas Boyd Fox and Evan Freed for their assistance with data collection and platform development, as well as Evan Forman, PhD, and Santiago Ontañón, PhD, for their contributions to the iterative study design.

Conflicts of Interest

None declared.

References


63. Arigo D, Torous J. Development of a mobile assessment tool for understanding social comparison processes among individuals with schizophrenia: two-phase survey study. JMIR Form Res 2022 May 02;6(5):e36541 [FREE Full text] [doi: 10.2196/36541] [Medline: 35499856]

Abbreviations

API: application programming interface

ICC: intraclass correlation coefficient

PA: physical activity

sr: semipartial correlation coefficient
How Reflective Automated e-Coaching Can Help Employees Improve Their Capacity for Resilience: Mixed Methods Study

Aniek Lentferink1, PhD; Hilbrand Oldenhuis2, PhD; Hugo Veltuijzen2, PhD; Lisette van Gemert-Pijnen1, Prof Dr

1Department of Psychology, Health & Technology, University of Twente, Enschede, Netherlands
2Marian van Os Centre for Entrepreneurship, Hanze University of Applied Sciences, Groningen, Netherlands

Corresponding Author:
Aniek Lentferink, PhD
Department of Psychology, Health & Technology
University of Twente
De Zul 10
Enschede, 7522 NJ
Netherlands
Phone: 31 505956217
Email: a.j.lentferink@utwente.nl

Abstract

Background: An eHealth tool that coaches employees through the process of reflection has the potential to support employees with moderate levels of stress to increase their capacity for resilience. Most eHealth tools that include self-tracking summarize the collected data for the users. However, users need to gain a deeper understanding of the data and decide upon the next step to take through self-reflection.

Objective: In this study, we aimed to examine the perceived effectiveness of the guidance offered by an automated e-Coach during employees’ self-reflection process in gaining insights into their situation and on their perceived stress and resilience capacities and the usefulness of the design elements of the e-Coach during this process.

Methods: Of the 28 participants, 14 (50%) completed the 6-week BringBalance program that allowed participants to perform reflection via four phases: identification, strategy generation, experimentation, and evaluation. Data collection consisted of log data, ecological momentary assessment (EMA) questionnaires for reflection provided by the e-Coach, in-depth interviews, and a pre- and posttest survey (including the Brief Resilience Scale and the Perceived Stress Scale). The posttest survey also asked about the utility of the elements of the e-Coach for reflection. A mixed methods approach was followed.

Results: Pre- and posttest scores on perceived stress and resilience were not much different among completers (no statistical test performed). The automated e-Coach did enable users to gain an understanding of factors that influenced their stress levels and capacity for resilience (identification phase) and to learn the principles of useful strategies to improve their capacity for resilience (strategy generation phase). Design elements of the e-Coach reduced the reflection process into smaller steps to re-evaluate situations and helped them to observe a trend (identification phase). However, users experienced difficulties integrating the chosen strategies into their daily life (experimentation phase). Moreover, the identified events related to stress and resilience were too specific through the guidance offered by the e-Coach (identification phase), and the events did not recur, which consequently left users unable to sufficiently practice (strategy generation phase), experiment (experimentation phase), and evaluate (evaluation phase) the techniques during meaningful events.

Conclusions: Participants were able to perform self-reflection under the guidance of the automated e-Coach, which often led toward gaining new insights. To improve the reflection process, more guidance should be offered by the e-Coach that would aid employees to identify events that recur in daily life. Future research could study the effects of the suggested improvements on the quality of reflection via an automated e-Coach.

(JMIR Hum Factors 2023;10:e34331) doi:10.2196/34331

KEYWORDS
self-reflection; stress management; resilience; eHealth; self-tracking; e-Coaching; mobile phone
Introduction

Background

Sustainable employability is, for a large part, negatively affected by stress, with one-third of work-related absenteeism among employees being caused by stress [1]. According to the European Union compass for action on mental health and well-being, more should be done in the preventive phase to increase employees’ capacity for resilience and reduce the risk of burnout [2]. Tackling stress at an early stage is vital because it can have negative consequences on health, well-being, and productivity [3]. To tackle stress at an early stage, it is necessary that employees cope effectively with the causes of the stress response (ie, stressors). Awareness about the stress response and the stressor is a prerequisite for employees to activate the desired behavior change, that is, effectively coping with the stressor. Moreover, employees also need to learn and select effective coping strategies to deal with the stressor [4]. Resilience is achieved when employees effectively deal with stress [5]. An employee’s capacity for resilience, defined as “the ability to bounce back after adversity” [6], is determined by the possession of several psychosocial and protective factors that influence the relationship between a stressor and the initial stress response. Examples of such factors are employees’ coping repertoires and emotion regulatory capacities [5].

Reflection is an important step in training employees’ capacity for resilience [5,7]. Reflection involves evaluating past experiences and learning from them with the aim of optimizing personal performance in future situations [8,9]. One of the ways in which reflection on stressful events improves resilience capacities is to prompt employees to search for ways to improve and adapt, recruit more coping strategies, and activate available resources such as social support or taking more time to complete a task [5]. It is useful to perform reflection soon after experiencing a situation that causes stress (reflection-in-action) and later (reflection-on-action) [10,11]. Stressful moments are opportune moments to perform a coping strategy, and a reassessment later in time can result in better recognition of stress or a stressor in future situations [5]. Another way in which reflection improves the capacity for resilience is that the negative event can be interpreted as less negative once time has passed, and individuals know the outcome of the stressor, which is often less severe than expected. This can lead to the situation being reframed into something more positive and unnecessary to worry about [12].

In traditional coaching settings, reflective coaching has received a great deal of attention as an effective and essential method to help coaches better understand and learn how to improve their situation [13,14]. The reflective coaching model [15], which is currently used in face-to-face coaching, includes four phases: (1) identification, (2) strategy generation, (3) experimentation, and (4) evaluation. The identification phase involves identifying issues worth solving and understanding why each of them is an issue; the strategy generation phase involves searching for and choosing possible solutions for the issue; the experimentation phase involves experimenting with the chosen strategies; and the evaluation phase involves evaluating the effectiveness of the strategy as a solution for the issue [15]. In short, reflection includes gaining awareness about the current situation and learning how to deal effectively with it or similar situations in the future.

Owing to the number of employees experiencing stress, labor-intensive face-to-face reflective coaching sessions to improve the capacity for resilience are not realistic [16]. eHealth technologies have the potential to coach users through the process of reflection without human involvement [17]. Self-tracking of stressful events and events related to resilience can result in awareness of the current situation [18]. Real-time measures of stress and resilience capacities (eg, heart rate variability) can be collected using self-tracking devices, such as smartwatches [19,20], or ecological momentary assessment (EMA) via smartphones. EMA “assesses individuals’ current experiences, behaviours, and moods as they occur in real-time and in their natural environment” [21].

eHealth tools that include self-tracking often present collected data in a graph for the user or as a simple summary. These persuasive technology features [22] can support users in observing their status and progress toward the desired behavior change [17]. However, previous research on self-tracking of health behavior indicates that awareness of one’s healthy lifestyle via self-tracking alone is not sufficient to effectuate the desired behavior change [18,23-25]. Through self-tracking alone, a great deal of the reflection process must be performed by the users themselves, such as gaining a deeper understanding of their current situation and deciding which coping strategy to apply. Cheo et al [23] stated that “the ultimate goal is to reflect upon one’s data, extract meaningful insights, and make positive change, which are the hardest part”. As described above, coaching during the reflection process performed by the user themselves is an effective and essential method to help employees extract meaningful insights and make positive changes [13,14]. End users and other stakeholders emphasized that coaching during reflection, in addition to the collection and summarization of data, was an important need for resilience training via eHealth technology [26].

Reflective automated e-Coaching has the potential to provide the necessary guidance that will aid in transforming awareness into behavior change. In this study, automated reflective e-Coaching is defined as supporting, advising, and guiding the user to evaluate past experiences and learn from these experiences for future improvement without the involvement of a human coach [9,27]. An automated e-Coach can personalize the coaching strategy based on self-tracking data and inputs from the user regarding their coaching needs, make use of persuasive features to motivate and stimulate behavior change [22], and be accessible 24/7 for users.

As we believe that reflective automated e-Coaching can affect behavior change, we aimed to study how employees using an automated reflective e-Coach perceive its effectiveness and usefulness. It is not only important to know the outcome of the guidance offered by the automated e-Coach, that is, its effectiveness, but also to gain an understanding of how the use of the different design elements of the automated e-coach and the interplay between them contributes to the outcomes, that is,
the usefulness of the design elements during reflection [28]. To the best of our knowledge, no previous study has evaluated this aspect. To date, few eHealth technologies combine self-tracking and e-coaching. These technologies offer personalized feedback and goal setting based on self-tracking data [29-31]. However, they do not offer support, advice, or guidance during reflection, which is what automated reflective e-coaching fully entails in our opinion. The results of the perceived effectiveness and usefulness of reflective automated e-coaching can lead to implications for future designs in the context of resilience training. To explore this, we developed a prototype of the BringBalance app, as described in the section below.

The research questions that we aim to answer are as follows:

1. According to employees, what is the perceived effectiveness of the guidance offered by the automated e-coach in the BringBalance app during their reflection on the self-tracking data and strategies to improve their capacities for resilience?
   - To what extent do employees gain insights into their current situation and strategies to cope effectively with current and future situations through the automated e-coach?
   - How did employees perceive their stress levels and capacity for resilience before and after using the automated e-Coach in the BringBalance app?

2. What is the usefulness of the design elements of the automated e-coach in the BringBalance app to guide reflection by employees on the self-tracking data and strategies to improve the capacity for resilience?
   - To what extent are the individual design elements of the automated e-coach in the BringBalance app and the interplay between these design elements, useful during the process of reflection by employees?
   - What stimulating and stagnating factors did employees experience during the use of the design elements of the automated e-coach in the BringBalance app during their reflection process?

The BringBalance App

The goal of BringBalance is to coach users through the process of reflection to strengthen their capacity for resilience. The app leads the user through the four phases of reflection from the reflective coaching model [15]. Each phase includes a set of modules in which users receive information via written text or videos and are asked to answer questions from the automated e-coach. Tools such as visualizations with summaries support the users in their reflection process. The BringBalance app is a product of “De Maar Training & Advies” and is based on their face-to-face coaching program, Working on Resilience [32]. Results from a pilot study on this face-to-face coaching program indicated positive effects on stress reduction [33]. In addition to the coaching program Working on Resilience, results from earlier studies on self-tracking and e-coaching for resilience training were also used during the design of the BringBalance app [26,27,34]. Other sources for creating the design of the BringBalance app were provided by the literature on reflection [10,12,15,24,35-37], coaching techniques [38-43], and persuasive design elements that can support the reflection process, such as visualization and personalization [17,22,44,45].

The prototype of the BringBalance app was created using The Incredible Intervention Machine (TIIM), a tool of the Behavioural, Management and Social Sciences (BMS) lab at the University of Twente that supports building and testing eHealth interventions [46]. The BringBalance program via the app took 6 weeks to complete. The design elements were offered to the user in Dutch through the BringBalance program in the TIIM app, and all the design elements together comprised the automated e-coach. See Figures 1 and 2 for screenshots of the selection of design elements and Table 1 for an overview of the content of the BringBalance program. The design elements are in italics in Table 1.

The reflective coaching model with its four phases [15] was translated into a format suitable for automated e-coach. During the identification phase (phase 1), the employee was stimulated to gain insights into situations (energy leaks and sources) related to stress and resilience to find opportunities for improvement via several EMA questionnaires. The term energy leak was chosen to indicate bodily responses to stress that activate the sympathetic nervous system, such as a quickened heart rate and breathing pace, resulting in lower physical levels of energy [19]. In addition, in the context of this study, energy leaks refer to situations that lead to low mental energy levels, that is, a feeling of mental exhaustion. The term energy source indicates those resources that activate the parasympathetic nervous system, lowering the heart rate and breathing pace, and are related to a higher level of mental energy. Energy sources can help one regain balance in one’s energy levels [47], that is, enable a person “to bounce back after adversity”—which is also the definition of the capacity for resilience [6,38].

Phase 2, the strategy generation phase, consisted of learning the six BringBalance techniques via short clips and training for the techniques a day later. These BringBalance techniques are based on exercises from the HeartMath Institute [48] and entail being attentive to one’s heart area and using one’s imagination to breathe in and out through it [33]. In addition, a heart rate variability (HRV) sensor (Inner Balance Trainer, HeartMath Institute), placed on the participant’s earlobe, provided the participant with biofeedback during the training. HRV biofeedback has been found to support self-regulation capacities [49]. The HRV indices enabled the participants to see any immediate effect of the technique on their HRV levels, which they could then use to adjust their performance. At the end of phase 2, users decided upon helpful strategies for their three most important energy leaks and energy sources with the help of the e-coach. These strategies could be BringBalance techniques, an energy source, or a self-chosen strategy [40].
Figure 1. Screenshots of the BringBalance program phase 1 and 2 in The Incredible Intervention Machine (TIIM) app, including a few ecological momentary assessment (EMA) questionnaires.

Figure 2. Screenshots of the BringBalance program phase 1 and 2 in The Incredible Intervention Machine (TIIM) app, including a few ecological momentary assessment (EMA) questionnaires.
In phase 3, the experimentation phase, users experimented with the chosen strategies and received reminders to do so at self-chosen moments [40]. After applying the strategies in real life, the users were asked to evaluate the strategy using an EMA questionnaire. In addition, the users were asked to report their energy balance every day. Energy balance was defined for participants as the balance between mental and physical energy-absorbing processes due to energy leaks and the processes that give them mental or physical energy from the energy sources [47].

All collected data from phase 3 were visualized in a graph and table and presented to the user in phase 4, the evaluation phase. In phase 4, the user evaluated whether the chosen strategies were the right strategies for their energy sources and leaks and whether their energy balance had improved. A more in-depth description of the BringBalance app, including the BringBalance techniques and persuasive design elements in BringBalance per phase of reflection, complying with the CONSORT (Consolidated Standards of Reporting Trials) guidelines for the reporting of eHealth interventions [50], can be found in Multimedia Appendix 1 [51,52].

### Methods

#### Participants

Participants were recruited via email sent to all employees by the Human Resources (HR) department of a software company with approximately 350 employees in the east of the Netherlands. The HR department informed potential participants about the objectives of the study, the BringBalance app, data collection and management, and the amount of effort requested for employee participation. Employees willing to participate were asked to fill in a web-based questionnaire with the validated Dutch version of the Perceived Stress Scale (PSS) [53-55] and an informed consent form. The inclusion criterion for participation was a score above 14 on the PSS, indicating a higher-than-average perceived level of stress [56,57]. This inclusion criterion was based on earlier studies performed by the authors [26,58], which showed that employees with a certain level of stress tend to have a higher motivation to complete the

<table>
<thead>
<tr>
<th>Phase</th>
<th>Duration</th>
<th>Requested from the user in this phase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase 1—identification</td>
<td>2 weeks</td>
<td>Three times during the week and once daily during the weekend: filling in the EnergyBalance questionnaire.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Once daily: reflecting on the collected data of the day before including the 4G scheme [38] asking the user to provide a more detailed description of the situation as well as their emotional state, physical state, cognitions, and behavior during the situation. The collected data from the day before were presented to the user via a table and graph.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>End of phase 1: choosing the 3 most important energy sources and leaks from a list with an overview of the collected energy sources and leaks.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Result: self-tracking data on the energy balance for comparison with phase 3, list of energy sources and leaks, and top 3 most important sources and leaks.</td>
</tr>
<tr>
<td>Phase 2—strategy generation</td>
<td>2 weeks</td>
<td>Every Monday, Wednesday, and Friday: learning 1 of the 6 BringBalance techniques.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The day after the introduction of the technique: receiving a reminder to practice the BringBalance technique with the inner balance trainer to obtain biofeedback during training the techniques.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>End of phase 2: Choosing strategies for their 3 most important energy sources and leaks. While selecting a strategy, users could receive guidance via the strategy database with an overview of all BringBalance techniques and tips for application in daily life or via e-coach’s guiding questions; setting implementation intentions [40] in which the strategies were linked to the energy sources and leaks and reminders with these implementation intentions on self-chosen moments for phase 3.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Result: strategies were chosen for the top 3 energy sources and leaks, implementation intentions were set including the strategies for the energy sources and leaks, and reminders were set with the implementation intentions.</td>
</tr>
<tr>
<td>Phase 3—experimenting</td>
<td>2 weeks</td>
<td>Daily: receiving reminders with the implementation intentions set in phase 2; experimenting with the chosen strategies (optional: using the Inner Balance sensor) according to the implementation intentions; evaluating the strategy for its effect with a strategy evaluation form after experimenting with a strategy, including questions on the effect of the strategy on mood and energy levels, and stimulators and demotivators; filling in the EnergyBalance questionnaire once daily.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Result: data on the evaluation of the strategies, self-tracking data on energy balance for comparison with phase 1.</td>
</tr>
<tr>
<td>Phase 4—evaluation</td>
<td>1 day</td>
<td>At the end of the program: receiving the data collected in phase 3 via visualizations in tables and graphs; evaluating if the strategies helped to prevent or resolve energy leaks and helped to make more use of energy sources; evaluating if the energy balance improved; advice on how to continue working on their energy balance after completion of the program.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Result: final reflection on the strategies and energy balance and advice on how to continue working on their energy balance.</td>
</tr>
</tbody>
</table>

*aThe design elements are shown in italics.*
intervention owing to a higher expected benefit compared with employees with lower stress levels. Finally, participants were required to own an Android (version 5.0 or higher) or iOS (version 10.0 or higher) smartphone.

A total of 45 participants filled in the questionnaire, with a response rate of 13%. Because 15 HRV sensors were available, 30 participants were invited to join either one of two sessions: November 2018 (n=15) or January 2019 (n=15). Participation in the study was voluntary.

Study Design, Data Collection, and Analyses

A convergent mixed methods design was applied “to obtain different but complementary data on the same topic” [59] for a more complete understanding of the problem [60]. The data collection included (1) a pre- and posttest survey, (2) EMA questionnaires in the BringBalance app, (3) log data, and (4) in-depth interviews. The pretest survey was completed before the BringBalance program; the EMA questionnaires and log data were collected during the BringBalance program; and the posttest survey and in-depth interviews were conducted after the BringBalance program. The collected data included perceived effectiveness (gaining insights [research question: RQ 1A], stress, and capacity for resilience [RQ 1B]) and perceived usefulness (utility of the design elements [RQ 2A] and stimulating and stagnating factors during the use of the design elements in the four phases of reflection [RQ 2 B]). The collected data included data on adherence to the intended use, dropout, app use, user motivation, usability, and experience with the BringBalance program in general. These data were used to confirm, explain, or nuance the results of the main outcomes of interest. Figure 3 shows a flowchart that includes an overview of the methods for data collection and integration of the data during collection and analysis. A data management plan was established according to the General Data Protection Regulation, a regulation for the protection of personal data inside and outside the European Union. More information on the data management plan can be found in the section Data Management.

Figure 3. Flowchart of methods for data collection and data integration. BRS: Brief Resilience Scale; EMA: ecological momentary assessment; O: Other data to explain or nuance results; PSS: Perceived Stress Scale; RQ: research question.

The Pretest Survey

The web-based pretest survey was completed using Qualtrics survey software (Qualtrics, Provo, UT) 7 to 1 day before the start of the BringBalance program. The pretest survey included (1) demographic characteristics (age, gender, function, and educational level), (2) the Dutch version of the PSS; range of possible scores: 0-40) [53-55], (3) the Dutch version of the Brief Resilience Scale (BRS; range of possible scores: 1-5) [6,61], and (4) ease of using a smartphone rated on a scale from 1 to 5. The latter question was self-developed and included as an indication of the participant’s smartphone skills. Both the he PSS [53-55] and BRS [6,61,62] are validated questionnaires. PSS was used to check whether the participants met the inclusion criteria. The pretest PSS and BRS scores were used to gain insights into the study population and to compare against posttest scores to assess perceived effectiveness on stress and resilience capacities (see the blue box in Figure 3). However, no causal effect of the guidance offered by the automated e-coach could be deduced from the study setup. Data from the pretest survey were uploaded to the SPSS (IBM Corp) to calculate descriptive statistics.

EMA Questionnaires in the BringBalance App

During the BringBalance program, participants were asked by the automated e-coach to complete several tasks throughout the reflection process. Participants were asked via a reminder on their smartphones to fill in EMA questionnaires related to a specific task. The app included 17 different EMA questionnaires spread over the four different phases of reflection, each with their own content, doses, and timing. Some EMA questionnaires were released at fixed moments during the BringBalance program.
program, whereas others were released based on a specific answer given in another EMA questionnaire. Multimedia Appendix 1 includes in-depth information on the setup of the EMA questionnaires, based on the reporting checklist from Van Berkel et al [63], along with examples of EMA questionnaires in the app. Figures 1 and 2 also include screenshots of the selection of available EMA questionnaires in the app. The answers to the EMA questionnaires provided insights into how users completed the reflection phases. These data provide insights into the perceived utility of design elements for reflection and stimulating and stagnating factors during reflection using the design elements. The last EMA questionnaire asked participants to report if they perceived a beneficial effect on their energy balance (yes or no) and if they had gained insights into their energy leaks and sources and strategies to improve their energy balance (yes or no). These data were used to determine whether participants gained insight into their current situation and strategies to improve their situation. Data were stored in the database of the BMS lab at the University of Twente and retrieved by uploading the data in Microsoft Excel files.

EMA data were used when it was necessary to further explore and interpret the results from the analyses of the interview data (see the green box in Figure 3). For example, when participants mentioned having difficulties interpreting a question from the automated e-coach, answers given on EMA questionnaires provided insights into the way users interpreted the question. In addition, EMA data were used as input for discussions during the interviews (see the orange box in Figure 3). For example, when a participant never used an element, it could say something about the perceived utility of the particular design element during the four phases of reflection. In addition, log data were more deeply analyzed when the posttest survey and interview data at the group level identified a result that needed to be explored further (see the green box in Figure 3). The data were stored in the database of the BMS lab at the University of Twente and could be retrieved in Excel files. Excel files were uploaded to SPSS, and descriptive statistics were calculated, such as frequencies and descriptive statistics were calculated, such as frequencies.

**Log Data**

Log data were collected via the TIIM app during the BringBalance program and included the following data for each participant: (1) which design element was completed, (2) the timestamp when the design element was delivered to the user, (3) the timestamp when the design element was returned by the user, and (4) the duration of completing the design element. Log data were used to confirm, explain, or nuance the results of the main outcomes of interest (perceived effectiveness and usefulness). First, log data were used to analyze adherence to the intended use and dropout. The intended use was set up by one of the researchers (AL) and was based on the minimum expected necessary use to be able to go through the phases of reflection. See Table 2 for the intended uses. Insights into adherence to the intended use and dropout were necessary to gain an understanding if the perceived effectiveness (perceived effect on stress, resilience capacities, or gaining insight) may have been affected by factors other than the design elements of the automated e-coach, such as lack of ease of use, user motivation, or personal reasons for nonadherence or dropout. Elaboration of reasons for nonadherence and dropout during interviews helped to explain the perceived effectiveness of the automated utility of design elements and stimulating and stagnating factors during the use of the design elements for reflection. Moreover, an overview of log data per participant was used during the interviews to discover the perceived utility and stimulating and stagnating factors during the use of different design elements (see the orange box in Figure 3). For example, when a participant never used an element, it could say something about the perceived utility of the particular design element during the four phases of reflection. In addition, log data were more deeply analyzed when the posttest survey and interview data at the group level identified a result that needed to be explored further (see the green box in Figure 3). The data were stored in the database of the BMS lab at the University of Twente and could be retrieved in Excel files. Excel files were uploaded to SPSS, and descriptive statistics were calculated, such as frequencies of adherence to the intended use per phase.

**Table 2.** Intended use for adherence.

<table>
<thead>
<tr>
<th>Phase of BringBalance</th>
<th>Intended use</th>
</tr>
</thead>
</table>
| Phase 1—identification | • The user completed 80% of the design elements “EnergyBalance” and “Reflection on the day before.”  
• The user finished the design element “Top 3 most important energy leaks and sources.” |
| Phase 2—strategy generation | • The user views 6 out of 7 (86%) short clips about strategies.  
• The user chooses strategies for at least 2 energy leaks and 2 energy sources.  
• The user sets implementation intentions for at least 2 energy leaks and 2 energy sources. |
| Phase 3—experimentation | • The user completes 80% of the EnergyBalance questionnaires.  
• The user completed at least 2 “strategy evaluation forms” per strategy. |
| Phase 4—evaluation | • The user evaluates 2 strategies for energy leaks and 2 strategies for sources.  
• The user evaluates the energy balance. |

*No absolute values can be provided because the number of received design elements varied between users.*
Posttest Survey

The participants were asked to fill in the web-based posttest survey via Qualtrics (Qualtrics) after they finished the BringBalance program and before they participated in the interviews. A total of 3 participants did not follow up on this because of time constraints. The full survey can be found in the Multimedia Appendix 2. The set of questions in the posttest web-based survey explored the following issues and was based on a survey used in an earlier study on the utility of persuasive design elements in an app for reflection [17]:

- The Perceived Stress Scale
- The Brief Resilience Scale
- Experience with BringBalance in general
- Motivation to complete BringBalance
- Perceived effect of the guidance offered by the automated e-coach on reflection outcomes: gaining insights into their energy balance and strategies to improve their energy balance
- The utility of the elements in the BringBalance app during the four phases of reflection

The participants reported their experience of using BringBalance in general by rating the BringBalance app on several aspects (scale of 1-10), such as usability, appeal, and integration into their daily life [64], as well as responding to three questions asking them to elaborate on their given ratings. In addition, the survey included two statements on their motivation to complete the BringBalance program. An example of a statement was “The BringBalance programme motivated me to reflect on my energy leaks and sources.” Insights into participants’ experiences with the technology and their motivation to use the technology were used to explain the underlying reasons for the perceived effectiveness and usefulness of reflection design elements [28,44].

Perceived effectiveness of the automated e-coach on reflection outcomes was measured in the posttest survey by three statements (5-point Likert scale from strongly disagree to strongly agree): (1) “The e-Couch has given me a clear overview of my most important energy leaks and energy sources”; (2) “Thanks to the e-Couch, I know what I could do in the future to prevent or resolve energy leaks”; and (3) “Thanks to the e-Couch, I know what I could do in the future to take more advantage of my energy sources.”

The main part of the survey consisted of questions regarding the experienced utility of the reflection design elements of the automated e-coach in the BringBalance app. Participants were asked to score the utility of each design element of the automated e-coach per phase of reflection that they received during the BringBalance program on a scale of 1 to 5. For example, “On a scale from 1–5, to what extent has the EnergyBalance questionnaire helped you gain insights into your energy leaks and energy sources? (1=not at all, 5=very much).” Each set of questions related to one phase of the BringBalance program ended with a blank space for participants to comment freely on the design elements of the automated e-coach for that specific phase.

The results of the posttest survey were used as inputs during the interviews (see the orange box in Figure 3). For example, the interviewee was asked to elaborate on the low scores given to the design elements of the automated e-coach. Posttest survey data were uploaded to SPSS, and descriptive statistics were reported for the group in total, which included the completers and noncompleters of the BringBalance program. Statistical analysis was not performed because of the small sample size (n=28).

In-Depth Interviews

Interviews were conducted one on one by the first author of this manuscript in person or via Skype (Microsoft) after the participants completed the BringBalance program. Recordings of the interviews lasted from 23 to 48 minutes. Furthermore, 7 of the 28 participants did not participate in the interviews because of practicalities.

In-depth interviews were held for confirmation and explanation, and to find nuances behind the answers given in the EMA questionnaires, the collected log data, and answers on the posttest survey (see the orange box in Figure 3). In addition, interviews were conducted to gain an understanding of the experiences, usability of the BringBalance app, perceived effectiveness of the automated e-coach, and how the process of reflection via the automated e-coach proceeded. The interview scheme was set up by the first author of this manuscript and finalized by all authors. The topics in the interview scheme were the user’s experiences in general, the usability of the app, reasons for nonadherence to the intended use and dropping out, the process of gaining insights into energy sources and leaks, related to the identification phase of the reflection process, and the process of gaining insights into when and what strategies to use, related to the strategy generation, experimentation, and evaluation phase of the reflection process. Subtopics for the reflection process included the design elements of the automated e-coach. The first 3 topics were discussed to obtain a sense of the experience with the app because experiences can affect the desired outcomes [28,65]. The elaboration of these topics by participants may also reveal the perceived utility of the design elements and stimulating and stagnating factors per phase of reflection. The latter 2 topics were discussed concerning the perceived effectiveness of reflection outcomes (ie, users’ insights into energy leaks and sources and strategies to improve their situation), the utility of the design elements of the automated e-coach and stimulating and stagnating factors during the use of the design elements. Results from EMA questionnaires, the posttest survey, and log data were used as inputs during the interviews (see the orange box in Figure 3). Participants were strongly encouraged to provide examples. The interview scheme can be found in Multimedia Appendix 3.

Interviews were transcribed verbatim. The transcripts were uploaded to Atlas.ti for qualitative data analysis. The coding scheme was created using inductive and deductive coding. Deductive codes came from the literature on reflection [15,35] and persuasive design elements [22] and included the design elements in the BringBalance app. Deductive codes for gaining insights from the participants were based on the level of

https://humanfactors.jmir.org/2023/1/e34331
reflection described by Durall et al [35]: no new insights, no reflection, recognition, and reflection. No new insights refer to insights that are a confirmation of what is already known, and recognition refers to quotes in which the user understands the data but acknowledges only what is expressed in the visualization of the data. Reflection involves gaining new insights via behaviors clearly associated with reflection, being surprised by the new insights, linking the insights to other experiences or situations in their daily life, or the insights affecting the beliefs or behavior of the user. No reflection refers to not obtaining any insights [35]. Open coding was performed for quotes that could not be labeled by deductive coding. Axial coding led to organizing codes into categories, removing synonyms, and splitting codes when necessary [66]. The initial coding scheme that resulted from coding the two transcripts was tested for intercoder consistency [67]. Two researchers (a student assistant, mentioned in the Acknowledgments section, and the first author of this manuscript) coded the two transcripts independently and discussed the differences until a consensus was reached. The discussions resulted in sharper descriptions of the codes. Finally, selective coding was performed to identify the themes that answered the research questions. During the process of selective coding, special attention was paid to finding contradictory quotes and differences between groups of participants, for example, between the study’s noncompleters and completers [66].

Mixing Strategies

Mixing strategies refer to those used to mix qualitative and quantitative strands [60]. All types of data collected were analyzed separately. As described above, some results from the analyses of one data source were inputted during the collection of another data source (eg, an overview of the log data per participant was used during interviews). Moreover, the results from different data sources per outcome of interest were compared to identify discrepancies and similarities between the results [60]. For example, the results on the utility of the design elements during reflection came from EMA questionnaires, posttest surveys, and interviews. This approach led to stronger evidence when similarities were observed and implications for further research when discrepancies were observed. Moreover, results from the analyses of one data source are often used to explain or nuance the results found during analyses of another data source.

Data Management

The data management plan was made in DMPonline (TU Delft, Delft) and in collaboration with experts on data management from the Department of BMS, University of Twente, to ensure that data collection and storage were performed according to the General Data Protection Regulation. Ethical approval was obtained from the Ethical Committee of the University of Twente (reference number: P-1531727676).

Results

Demographic Characteristics

A total of 28 participants started using the BringBalance program, of which 21 (75%) were men and 7 (25%) were women, with a total average age of 36.5 (SD 9.7) years. Average PSS scores were 16.8 (SD 5.0) and BRS scores were 2.9 (SD 0.8). Table 3 provides an overview of the participants’ demographic characteristics.

Table 3. Demographic characteristics of the participants and their ease of using a smartphone.

<table>
<thead>
<tr>
<th></th>
<th>Study noncompleters (n=14)</th>
<th>Completers (n=14)</th>
<th>Total (n=28)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Man</td>
<td>8 (57)</td>
<td>13 (92)</td>
<td>21 (75)</td>
</tr>
<tr>
<td>Woman</td>
<td>6 (43)</td>
<td>1 (7)</td>
<td>7 (25)</td>
</tr>
<tr>
<td>Nonbinary</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Age (years), mean (SD)</strong></td>
<td>37.4 (11.2)</td>
<td>35.6 (8.3)</td>
<td>36.5 (9.7)</td>
</tr>
<tr>
<td><strong>Educational level, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>University of applied sciences</td>
<td>10 (71)</td>
<td>8 (57)</td>
<td>18 (64)</td>
</tr>
<tr>
<td>University</td>
<td>4 (29)</td>
<td>6 (43)</td>
<td>10 (36)</td>
</tr>
<tr>
<td><strong>Perceived Stress Scale score, mean (SD)</strong></td>
<td>16.4 (4.9)</td>
<td>17.1 (5.2)</td>
<td>16.8 (5.0)</td>
</tr>
<tr>
<td><strong>Brief Resilience Scale score, mean (SD)</strong></td>
<td>3.2 (0.8)</td>
<td>2.7 (0.7)</td>
<td>2.9 (0.8)</td>
</tr>
<tr>
<td><strong>Ease of using a smartphone, mean (SD)</strong></td>
<td>4.6 (0.5)</td>
<td>4.6 (0.5)</td>
<td>4.6 (0.5)</td>
</tr>
</tbody>
</table>

Range of possible scores is 0 to 40.

**Characteristics of Participants Not Taking Part in Interviews**

Of the 28 participants, 7 (25%) did not participate in the interviews because of practicalities. Of these, 5 participants dropped out, of which 1 participant adhered to the intended use until dropping out. Other dropouts did not adhere to the intended use during all phases. The remaining 2 participants completed the BringBalance program and adhered to the intended use in phase 2. The average PSS score of the participants who did not...
participate in the interviews was 17.9 (SD 3.0), and the average BRS score was 3.0 (SD 0.6).

**Adherence and Dropout**

The log data indicated that none of the 28 participants adhered to the intended use, mainly due to an adherence rate of 0% in phase 3 ([Figure 4](#)). The adherence rates for the remaining phases were 25% (n=7) in phase 1, 50% (n=14) in phase 2, and 21% (n=6) in phase 4. The lowest adherence score in phase 3—experimentation—can be explained via interview data by a loss of overview by participants or their low-quality input in the earlier steps of the reflection process. According to the participants, the latter was a result of the guidance by the e-coach that steered them in a direction that was too specific (described in further detail in sections phase 1—identification and phase 2—strategy generation), lack of available time by participants, or the low priority given to the app. See [Figure 4](#) for adherence rates among completers, study noncompleters, and the total group of participants.

A total of 14 participants completed the BringBalance program. Most participants dropped out in phase 2 (n=11, 39%). From the interview data and reports via email, the primary reason for dropping out was the program’s difficult integration into the daily life of participants owing to their full schedule (n=5, 18%), followed by the e-coach requiring too much of their attention and time (n=3, 11%), personal circumstances (n=3, 11%), or loss of interest in the program (n=3, 11%).

[Figure 4. Adherence to intended use.](#)
Motivation to Complete and Ratings of the BringBalance Program

On average, participants rated the BringBalance program 6.5 out of a score of 10 (SD 1.0) in the posttest survey. On average, the BringBalance program scored 7.4 (SD 1.1) on being informative, 5.3 (SD 1.5) on usability, and 4.3 (SD 1.4) on integration in daily life (scale 1-10). Participants rated their motivation to reflect on energy sources and leaks as 3.3 (SD 1.0) and motivation to reflect on strategies as 2.7 (SD 0.7; scale 1-5). See Table 4 for an overview of participants’ ratings of the BringBalance program in general and their motivation to complete the BringBalance program, as determined by their scores in the posttest survey.

Table 4. Results of the posttest survey on participants’ ratings of the BringBalance program in general, their perceived effectiveness of the e-coach in the BringBalance program, and their motivation to complete the program.

<table>
<thead>
<tr>
<th>Questiona</th>
<th>Study noncompleters (n=14)</th>
<th>Completers (n=14)</th>
<th>Total (n=28)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BringBalance app in general (scale 1-10), n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Score BringBalance in general, mean (SD)</td>
<td>6.2 (0.8)</td>
<td>6.9 (1.2)</td>
<td>6.5 (1.0)</td>
</tr>
<tr>
<td>The appeal of the content of the app, mean (SD)</td>
<td>6.8 (1.2)</td>
<td>6.9 (1.0)</td>
<td>6.9 (1.1)</td>
</tr>
<tr>
<td>Perceived utility of the app, mean (SD)</td>
<td>5.6 (1.5)</td>
<td>7 (1.0)</td>
<td>6.3 (1.4)</td>
</tr>
<tr>
<td>Usability of the app, mean (SD)</td>
<td>5.5 (1.3)</td>
<td>5.1 (1.6)</td>
<td>5.3 (1.5)</td>
</tr>
<tr>
<td>Integration in daily life, mean (SD)</td>
<td>3.6 (1.5)</td>
<td>4.9 (1.1)</td>
<td>4.3 (1.4)</td>
</tr>
<tr>
<td>Informative, mean (SD)</td>
<td>6.7 (1.4)</td>
<td>7.6 (0.8)</td>
<td>7.4 (1.1)b</td>
</tr>
<tr>
<td>Advise the app to a colleague, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>6 (43)</td>
<td>10 (71)</td>
<td>16 (57)</td>
</tr>
<tr>
<td>No</td>
<td>8 (57)</td>
<td>4 (29)</td>
<td>12 (43)</td>
</tr>
<tr>
<td>BringBalance met my expectations, mean (SD)</td>
<td>5.4 (1.7)</td>
<td>7.1 (1.1)</td>
<td>6.3 (1.6)</td>
</tr>
<tr>
<td><strong>Motivation: (1=strongly disagree, 3=neutral, 5=strongly agree), mean (SD)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The BringBalance program motivated me to reflect on my energy leaks and sources.</td>
<td>2.9 (0.9)c</td>
<td>3.8 (1.0)</td>
<td>3.3 (1.0)</td>
</tr>
<tr>
<td>The BringBalance program motivated me to reflect on my chosen strategies for my energy leaks and sources.</td>
<td>2.4 (0.5)c</td>
<td>3.1 (0.7)</td>
<td>2.7 (0.7)</td>
</tr>
</tbody>
</table>

a Abbreviations of the questions were used in the table.
b n=20.
c n=13.

Perceived Effectiveness of Reflection

Among completers, pre- and postintervention scores on the PSS were mean 17.1 (SD 5.2) before and mean 16.9 (SD 3.5) afterward and the scores on the BRS were mean 2.7 (SD 0.7) before and mean 2.9 (SD 0.6) afterward. Of the 14 completers, 10 reported in the last EMA questionnaire that they had improved their energy balance. The remaining 4 completers reported that they gained insight into their energy balance. The interview data showed that participants gained insights into their energy sources and leaks (level of reflection: reflection). An example of a quote that shows reflection is the following: “My girlfriend is quite outgoing and she likes to constantly do things. I noticed that costs me a lot of energy and I did too little to recharge by tucking away in my own little world for half an hour and then just enjoy the social things again” (respondent #6, study completer). Some participants wondered whether they were on the right track with their reflective process (level of reflection: recognition). “Am I now thinking in the wrong direction or do I make it bigger than it actually is?” (respondent #1, study noncompleter). The interview data demonstrated that many participants gained insights into adaptive coping strategies and had an idea of when to use these techniques in daily life. However, the actual integration of the techniques in daily life was experienced as challenging by participants because of difficulties in learning the techniques (level of reflection: recognition). Elaboration on the difficulties encountered during this integration process are described in section Phase 2—Strategy Generation.

The Usefulness of Design Elements During Reflection

Overview

Multimedia Appendix 4 includes the scores of each design element of the automated e-coach for its utility for reflection per phase of reflection. Utility scores (scale 1-5) are described below for elements that were discussed intensively by users during the interviews, indicating that these elements were evoked a lot among the users. Multimedia Appendix 5 provides an overview of the identified themes, that is, the stimulators and stagnating factors according to the participants, per phase of reflection, and the specific design elements of the e-coach. The most important ones, that is, mentioned by many respondents or those with a great impact on reflection outcomes, per phase of reflection are described
Phase 1—Identification

According to the participants, the first phase of the BringBalance app was easy to complete independently. Most participants mentioned that they were able to gain an understanding of their energy balance during their reflection, as guided by the e-coach. Based on the interview data, an understanding of their energy balance was mostly obtained via the list with collected sources and leaks at the end of phase 1, in which the most common sources and leaks were often perceived as the most important sources of their energy balance. This design element received a mean utility score of 3.7 (SD 0.8). Contextual information about the situation related to the energy sources or leaks was necessary to reconstruct the situation from the previous day, especially when the user’s data showed little variance. The table with a visualization of the data collected from the previous day received a mean score of 3.6 (SD 0.8) and the graph a 2.9 (SD 1.1) in the posttest questionnaire.

During the interviews, 5 participants mentioned that the 4G scheme questions were superfluous. Another group of participants found the element useful. The 4G scheme included questions asking the user to provide a more detailed description of the situation, as well as their emotional state, physical state, cognition, and behavior during the situation. The average utility score of the 4G scheme was 3.3 (SD 1.0). Participants who found it useful described that reflection later in time led to the observation of more relevant aspects than reflection in close occurrence to the situation. In addition, the participants experienced that the questions stimulated an in-depth reflection on the source or leak. Moreover, 4 participants had difficulties in recognizing indicators for energy sources and leaks and therefore with filling in the 4G scheme questions. Some participants experienced that filling in the 4G scheme questions made them understand their indicators for energy sources and leaks and enabled them to be better indicate an energy source or leak in future situations. “Over time, you become more and more aware that your body reacts in a certain way” (respondent #17, study noncompleter).

A total of 3 participants mentioned that the guidance offered by the e-coach led to the identification of sources and leaks that were too specific. “The tool only focuses on such a micro-moment, and it will not zoom out to a category or something” (respondent #21, completer). Some participants believed that they could have gained a higher level of reflection if they had reflected on their self-tracking data in dialogue with another person.

Phase 2—Strategy Generation

According to the completers, the design elements for learning the BringBalance techniques, including short clips, were perceived as helpful in the process of understanding when and what strategies to use. During the interviews, the users mentioned that they were able to learn the principles of the techniques.

Reflected in the interview data, practicing the techniques were perceived as a crucial part of understanding which techniques are useful for their situation. However, practicing the techniques in daily life was experienced as somewhat difficult without the presence of a relevant situation in which the technique might be useful. “Usually, the conditions were not right for the technique to work. I would call it ‘dry swimming.’ [...] Then you rush practicing the technique and you don’t really practice anymore” (respondent #16, completer). Being attentive to indicators of sources and leaks, identified through the 4G scheme in phase 1, was mentioned by a few as a prerequisite to understanding when to apply the techniques in daily life. To master these techniques, many perceived 2 weeks as too short a time.

Of the participants that used the Inner Balance Trainer, 65% (11/17) reported in the posttest survey that they found it useful to receive HRV biofeedback practicing strategies. It convinced participants regularly of the potential effect of the technique on physiological stress reactions in future stressful situations, which was reflected in the interview data. Some participants had difficulties interpreting the results, were uncertain when to perform the measurements with the sensor or saw no change in scores before and during practicing as the scores were indicated as good from the start.

Often, users mentioned that connecting strategies to the most important leaks and sources stimulated their mental processes on how to integrate the techniques into their daily lives. Most participants said that they were able to choose strategies using the tools in the app. Log data showed that 11 people chose the strategy database, with an overview of the BringBalance techniques and their tips for application in daily life, as a tool to help them decide on a strategy and gave this element a mean utility score of 3.6 (SD 0.8). In addition, 5 people chose the e-coach’s guiding questions and gave this element a mean score of 4.7 (SD 0.6). One participant remarked on this specific tool help via the e-Coach: “Those questions helped to think a bit more towards a certain direction. That made me think: ‘What was my energy leak about?’ And based on that, I started searching for a technique in that direction” (respondent #6, completer). A few participants expressed doubts if the strategies they had chosen were the right ones for their sources and leaks.

The participants mentioned that poor input from previous phases made it difficult to decide on strategies that were sometimes irrelevant. Three participants mentioned that the identified leaks and sources were no longer relevant and 6 participants mentioned that they did not master the techniques in this phase. In addition, participants missed discussing this step with someone else who might have helped them determine whether they had made the right choice or advised them about other possible options.

Most participants found that the element, setting implementation intentions, in which the strategies were linked to the energy sources and leaks, stimulated their intention and mental process to integrate the techniques into their daily lives, although some found that the element steered them too much toward goals that were too specific.
**Phase 3—Experimentation**

The interviews revealed that many participants experienced difficulties during the experimental phase. Although the steps were experienced as logical in theory, they mentioned that leaks and sources did not recur anymore during phase 3 and that its duration was too short to experiment. “It is very difficult to get there within a week and a half. […] You ask yourself, did that technique help? And you don’t know for sure, and think: Maybe it was only a coincidence that the conversation went a little better” (respondent #24, study noncompleter). Log data showed that many started this phase later than planned, leaving little room for experimentation.

On average, personally set reminders along with the set implementation intentions scored 2.4 (SD 0.8) on utility. During the interviews, participants mentioned that reminders related to leaks and sources that occurred randomly over time did not trigger their application of a strategy, as the reminders were not “just-in-time.”

The evaluation of strategies began in phase 3 by evaluating every moment they performed a strategy with the strategy evaluation form (utility score on the posttest survey: mean 2.7, SD 0.8). Some participants experienced these forms as too repetitive and generic. “I can imagine that with the Zsleep or Flex technique, different questions come in handy” (respondent #6, completer). Depending on the specific strategy or situation, some participants said that they did not find it necessary to complete the strategy evaluation form each time. For others, the evaluation forms were a trigger to start the evaluation process.

**Phase 4—Evaluation**

Half of the participants who went through the elements of evaluation acknowledged the utility of evaluating strategies as a conclusion of the BringBalance program. However, almost all participants mentioned that they had insufficient data collected in phase 3 to perform a comprehensive evaluation of the strategies and their energy balance. Participants would have filled in more strategy evaluation forms in phase 3 if they knew in advance that they would later receive the collected data of these strategy evaluations as visualizations of the collected data in a table and graph from phase 3.

**Discussion**

**Principal Findings**

To improve the capacity for resilience through self-reflection, this study’s main aim was to examine the perceived effectiveness of the guidance offered by the automated e-coach in the BringBalance app during the reflection process on stress and resilience capacities among employees. In addition, this study’s goal was to determine the usefulness of the design elements of the automated e-coach for reflection, and the stimulating and stagnating factors during the use of the design elements.

**Perceived Effectiveness**

Pre- and posttest scores on perceived stress and resilience capacities were not significantly different among the completers of the BringBalance program. It should be noted that no statistical tests were performed because of the small sample size. Most completers reported an improved energy balance and insights into their principal energy leaks and sources as well as effective strategies for improving their situation. The reflection outcome of “linking these insights to other experiences or situations” by integrating the techniques into their daily lives was often not achieved.

**The Usefulness of Design Elements for Reflection**

Participants were easily able to reflect on self-tracking data and decide their most important energy leaks and sources with the design elements of the e-coach. Participants experienced difficulty integrating strategies relevant to their energy leaks and sources into their daily lives and reflecting on whether their chosen strategies were the right ones with the design elements of the e-coach.

Important stimulators for the process of reflection on self-tracking data were the design elements of the automated e-coach that stimulated the re-evaluation of situations and the observation of trends in the collected data through the breakdown of the reflection process into smaller steps and visualizations, including visualizations of the data via a table with an overview of sources and leaks from the previous day, and a list of sources and leaks at the end of phase 1. Some participants experienced that the re-evaluation later in time led to the ability to gain a larger perspective, leading to their understanding of more relevant details of a situation. In addition, contextual information added to the visualizations about the situation related to the energy source or leaks was necessary to be able to re-evaluate the situation later in time.

A stagnating factor for some participants was that the guidance offered by the e-coach led to the identification of sources and leaks that were too specific. Although most participants found it easy, some had difficulties recognizing physiological, mental, and emotional indicators of sources and leaks. These indicators are required in the 4G scheme.

The important stimulators for the process of reflection on strategies were (1) the short clips in which the participants learned the principles of the techniques, (2) the heart rate variability biofeedback to help them understand the principles and stimulate the effect of the BringBalance techniques on physiological stress reactions, (3) design elements that stimulated practicing the techniques because this rehearsal was perceived as a crucial step in the reflection process, and (4) the tools to help them decide upon the strategies and set up implementation intentions as these elements stimulate the user’s mental process on how to integrate the strategies into their daily lives. Participants found it useful to link the strategies to the sources and leaks, although, in practice, this did not bring about the desired results.

The most important stagnating factor for this lack of success was the low-quality input from previous steps in the process, such as the very specific energy sources and leaks identified in phase 1. The design elements to set up implementation intentions and reminders tended to lead participants excessively toward a specific context in which the strategy should be performed. In practice, this left little room for experimentation as the situation...
often did not recur. In addition, many participants experienced a lack of mastering the techniques in their daily lives owing to perceived time constraints, lack of relevant situations in which to practice, and doubts about performing the techniques in the right manner. These factors led to little experimentation and data collection in phase 3 and, therefore, difficulties for evaluation in phase 4, which involved answering the question whether the strategies were the right ones for dealing with the participant’s energy sources and leaks.

Comparison With the Literature

Perceived Effectiveness

In contrast to this study’s results indicating that scores on perceived stress and resilience were not much different in pre- and posttest scores, Rijken et al. [33] observed a tendency toward improvement in stress-related outcomes for the face-to-face program on which BringBalance was partly based. It should be noted that no statistical analysis could be performed in the BringBalance study owing to the small sample size. In addition, the prototype of the BringBalance program used in this study scored rather low on usability and integration in daily life, which likely affected the effectiveness of the guidance offered by the automated e-coach in the BringBalance app during the reflection process [65,68]. Still, there is a possibility that the element of reflection via human dialogue has played a role in the differences observed in the effectiveness of stress measures between the results of Rijken et al. [33] and this study, as this element was an important difference between the two programs. Some participants also mentioned the potentially stimulating role of human dialogue during reflections. The elaboration on how to deal with this issue in future designs is further discussed below.

The Usefulness of the Design Elements for Reflection

An important stimulating factor in the reflection process guided by e-coach seems to be the breakdown of the reflection process into smaller steps. These steps seemed to trigger participants to rethink their situations, which led to the observation of trends and a deeper understanding of their indicators of stress and resilience. The same process was observed in a study by Isaacs et al. [12], who found that participants defined as recorders (those who reported the event once) and those defined as reflectors (those who reflected on the event multiple times), both benefitted from their reflections, although reflectors were more likely to observe patterns and learn from these events to improve future performance.

Three important stagnating factors during the reflection process were (1) difficulties participants had in recognizing indicators for the presence of energy sources and leaks; (2) the identification of specific energy sources, leaks, and implementation intentions as guided by the e-coach; and (3) a perceived lack of availability. Although these stagnating factors were not experienced by all participants, targeting them can significantly impact reflection outcomes in a positive way for participants who experienced them.

First, participants in this study elaborated on the positive effect of being consciously aware of physiological, mental, and emotional indicators for their sources and leaks, including (1) being better able to recognize the presence of a source or leak in the future and (2) to identify opportunities for applying a strategy, known as “trigger identification” in the Systematic Self-Reflection Model of Resilience. This model emphasizes the importance of self-reflection in the process toward resilience [5]. Moreover, reflection on cognitions and emotions can help explain the behavior of the participant in a situation of interest and can lead to a higher level of understanding of their situation [38]. However, some participants in this study were not consciously aware of their indicators during the situation, that is, reflection-in-action, or found it difficult to reproduce the physiological, mental, emotional, and behavioral indicators concerning the situation when it occurred the next day, that is, reflection-on-action. This difficulty can negatively impact their reflection outcomes [38].

To identify the indicators effectively, both reflection-in-action and reflection-on-action are important [10,11]. Difficulties with reflection-in-action can be the employee’s limited ability to reflect under high levels of stress [5] or the concept of alexithymia because not everyone can recognize emotional responses [69]. Alexithymia can also explain difficulties with reflection-on-action because attention increases the likelihood of recalling the situation later [70]. In addition, other factors that negatively affect recall can explain difficulties with reflection-on-action, such as motivation and fatigue [71]. Proper guidance during reflection-in-action can solve problems with reflection-on-action, and vice versa. For example, problems due to alexithymia or recall may be solved by notifying the user just in time about the presence of stress symptoms and stimulating them to pay conscious attention to triggers in action [72]. Moreover, as mentioned by the participants, contextual information is necessary to recall the situation a day later and making notes in close occurrence is one method that seems to effectively tackle recall problems [10]. In this study, reflection-on-action was perceived as useful by participants because it enabled them to observe more details later in time. Reflection-on-action can also positively affect one’s overall reflection as the initial intensive stress response is diminished [5].

Second, the automated e-coach in the BringBalance app stimulated the participants’ intention to do something about the situation. However, a loss of relevance to continue behavior change was experienced when the identified sources and leaks were too specific. The problem of the limited applicability of previously collected data on well-being to current situations has been observed more often [24]. One way to maximize the applicability of specific situations to current situations might be to start choosing a strategy based on the underlying values and personal goals of the identified sources and leaks [5]. Situations that involve a mismatch between the current coping strategy and personal values and goals increase the need to do something about the situation [5]. Therefore, the underlying goals and values may serve as trigger points for adaptive coping strategies. The increased chances of recurrence can also lead to more opportunities to practice the techniques, which was mentioned by participants as a crucial step and is acknowledged in the literature as well: “The strengthening of resilience is a
process of experiential learning and more specifically learning through reflection on doing” [5].

Finally, it is unlikely that participants were constrained by the actual time needed to interface with the BringBalance program, which was approximately 15 minutes per day. This response was more likely a result of perceived time constraints caused by their busy schedules. This conclusion is based on the low scores given by the participants regarding the integration of the BringBalance program into their daily lives. Moreover, in the context of work, it seems that employees prioritize work-related activities over resilience [58] or learning via reflection [73].

Strengths and Limitations

First, our study population consisted of participants with high educational levels. As reflection relies on the analytical skills of a participant [74], it might be that the performance during the reflection process and the need for guidance from the automated e-coach by our study participants are different for the overall working population.

Second, the sample size was too small to conduct statistical analysis of pre- and posttest scores on perceived stress and resilience capacities and differences in scores given by study noncompleters and completers on the utility of design elements. This limitation restricted the strength of some of our conclusions. Although the statistical power was low, this study’s results did meet the primary aim of this study, namely to explore the potential of guidance offered by an automated e-coach during the participant’s reflection process for resilience training and to ascertain implications for future designs based on the results; therefore, valuable insights that can support future design were obtained.

Third, although low adherence rates are common for prototype versions of eHealth technology, none of the participants precisely adhered to the intended use. On the one hand, low adherence distracted the user from the original goal of the program, which was to reflect on improving resilience capacities, and likely affected the effectiveness. On the other hand, reasons for low adherence revealed important stagnating factors for reflection guided by the automated e-coach, such as the loss of relevance to continue owing to specific energy leaks and sources. Moreover, it should be noted that the setup of adherence to the intended use was based on 1 researcher’s expectations of the minimum necessary use by the user. This expectation may have been too ambitious as no participant adhered to the intended use, and results indicated that most participants gained insights into their energy leaks, sources, and strategies to improve their energy balance.

Finally, 7 participants were not involved in the interviews owing to practicalities, which might have affected the validity of the qualitative results. A relatively higher number of study noncompleters were observed among noninterviewees, and the noninterviewees’ PSS scores tended to be somewhat higher in comparison with the interviewees, although statistical tests could now be performed. However, similar characteristics were observed among interviewees, as 9 were study noncompleters and 7 scored higher than the average PSS score of the noninterviewees. This suggests that the validity of the qualitative results was not significantly influenced to a large extent.

Regarding the strengths of this study, the first is that the BringBalance program’s design was strongly based on literature and was created in close collaboration with stakeholders. These 2 aspects increase the chances of improving uptake and creating an impact on eHealth technology [65]. The participants perceived the design decisions made for the content in the app as logical and interesting. The usability of the app and its integration into daily life are points of attention. This can be explained by the limited options in the way the prototype could be developed. Usability and integration issues can be overcome when an app is developed with a higher level of fidelity [68].

Second, a mixed methods approach was used in this study. Results from one data source were used during the collection of another data source (eg, log data were used as inputs during interviews) or results from one data source were a trigger to explore more profoundly into the data from another source (eg, to review the log data to explain the lower scores given on the guidance offered by the e-coach by study noncompleters in comparison with completers). This enabled us to confirm or question the results of one approach to another. In addition, it enabled a deeper interpretation of the results by finding nuances in the data from other approaches.

Implications for Future Design and Research

To our knowledge, this study is the first to provide insights into the design elements of an automated e-coach that can simulate the self-reflection process, from the identification of relevant events to the evaluation of strategies [15], without support from a human coach. Future design and research can begin by focusing on the effects of making more and better use of persuasive features during the reflective automated e-coaching process, based on the 3 stagnating factors described above. Persuasive features can stimulate users’ motivation for behavior change and are shown in italics in the discussion below [22].

First, as described above, trigger identification is an important aspect of the reflection process and can result in both reflection-in-action and reflection-on-action. Continuous biofeedback, a form of persuasive self-tracking feature [22], creates a unique opportunity to receive timely external feedback [75] when stress is present. Moreover, biofeedback can be used to determine when the intensity of the stress response is diminished to some extent, which could have a positive effect on the quality of one’s reflection [5]. Several commercially available wearable devices are capable of continuous measurement of the physiological responses related to stress and resilience capacities, such as HRV measurements [19,76]. These measures can indicate within minutes that stress is present or when stress is decreased and, hence, signal to employees their capacity for resilience [76].

Second, the automated reflective e-coach should offer guidance in translating specific events into overarching goals and values that recur in daily life. The e-coach can help the user split the complex behavior into a higher perspective that oversees the collected data and breaks it into short and simple tasks, which is related to persuasive feature reduction [22]. For example, the
e-coach can ask the user to answer additional questions regarding their underlying opinions, values, qualities, and drivers to learn and understand their goals and values in daily life [74]. This implication for future design can also improve the technology’s effectiveness on the desired behavior change and user motivation, as this way the content better suits the user’s context of working and living [22,26].

Third, we propose a more dynamic process in which users can decide the pace of completing a phase, related to the persuasive feature of personalization, thereby avoiding poor inputs from previous steps in the reflection process owing to perceived limitations in time.

Finally, some participants believed that reflection in a dialogue with another person would lead to higher levels of reflection. This dialogue was also desired by participants to eliminate personal doubts about the individual reflection process. The involvement of a professional coach limits the program's scalability. Therefore, the first implication is that peer groups within organizations could facilitate dialogue. These peer groups can be organized according to the persuasive features of social facilitation. Using this feature, users can contact peers through the app [22]. Previous literature has found that peer guidance during reflective practices improves the reflective process [77,78].

Second, automated e-coaches could match human-to-human dialogue to a greater extent. This technological development is currently on the research agenda [79,80]. To match a human-to-human dialogue, the e-coach should have high surface credibility via fluent dialogue, and the user must experience the e-coach as a real human, an achievement that still requires considerable research and testing. However, some persuasive features that are rather easy to implement can improve the surface credibility of the currently available automated e-coaches by applying a high level of personalization, for example, by regularly selecting coaching messages based on previous inputs given by the user or repeating these inputs in messages, and the e-coach should adopt a social role, for example by greeting the user by name [22,79,80].

A follow-up study using an updated prototype of higher fidelity, including these aspects, can be performed to test the effects of the guidance offered by the automated e-coach on stress and resilience capacities, and gaining of insights on a larger scale, also including employees with lower educational levels. Again, a mixed methods approach should be applied to study both the effectiveness of the automated e-coach on stress, resilience, and reflection outcomes and to understand which design elements contribute to the effectiveness and why. Specifically, future research can, for example, combine log data of the continuous biofeedback (eg, when the e-coach offers guidance to perform reflection-in-action and reflection-on-action) with the participant’s answers to the EMA questionnaires to study the output of the reflection process during moments that are in close occurrence to the stressful situation and during moderate levels of stress.

Conclusions
The results of this study provide insights into the potential of automated e-coaching to guide employees during the reflection process for the purpose of resilience training. Most completers reported an improved energy balance and insights into their principal energy leaks and sources as well as effective strategies for improving their situation. The results indicate that an automated e-coach can guide employees during the reflection process on self-tracking data toward a deeper understanding of their situation and possible strategies to improve their situation. Design elements that stimulated the re-evaluation of situations and observation of trends stimulated the reflection process. It was more difficult to guide the employees via an automated e-coach to integrate the strategies into daily lives and reflect on whether the chosen strategies were the right ones. Future designs of the automated e-coach should make more and better use of persuasive features to support and motivate behavior change. Future research should focus on testing the effects on the reflection process by equipping the automated e-coach with more and improved persuasive features, as suggested above.

Acknowledgments
The authors wish to thank Ewold de Maar, owner of the De Maar Training and Advies, for his help and enthusiasm during the preparation of this study. Together with Ewold, the BringBalance program was developed. His original coaching program “Working on Resilience” (Werken aan veerkracht) was used as a basis for the BringBalance program and was translated together with AL into a smartphone version including reflective automated e-coaching. In addition, the authors would like to thank the Behavioural, Management and Social Sciences (BMS) lab of the University of Twente for their support during the development of the prototype version of BringBalance using The Incredible Intervention Machine application, and the consultation and problem-solving services offered during the execution of the study. In addition, the authors extend their gratitude to Anouk Burgler, who offered help during the execution of the study as part of her internship. Last but not least, many thanks to the Human Resources (HR) department of the organization where the study took place for their help during the recruitment of participants, and special thanks to all the participants who participated in this study.

This study was partially funded by Menzis. Menzis was not involved in the study design, execution, or reporting. De Maar Training and Advies were mainly involved in the design of the BringBalance Program and had limited involvement in the design of the study. During the study’s design, some questions were added to the posttest survey in which De Maar Training and Advies were especially interested in deciding if the investment into a full version of the BringBalance app might be worthwhile.
Data Availability

The data sets used and analyzed during this study are not publicly available because no consent was provided by the respondents before data collection took place. The data sets are available from the corresponding author on request.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Description of the BringBalance app according to the consort guideline on reporting eHealth.

[DOCX File, 595 KB - humanfactors_v10i1e34331_app1.docx]

Multimedia Appendix 2

Post-test survey.

[DOCX File, 45 KB - humanfactors_v10i1e34331_app2.docx]

Multimedia Appendix 3

Interview scheme BringBalance.

[DOCX File, 20 KB - humanfactors_v10i1e34331_app3.docx]

Multimedia Appendix 4

Utility scores for the design elements of BringBalance per phase of reflection.

[DOCX File, 22 KB - humanfactors_v10i1e34331_app4.docx]

Multimedia Appendix 5

Summary table of stimulators and stagnating factors for reflection per phase of reflection.

[DOCX File, 26 KB - humanfactors_v10i1e34331_app5.docx]

References

32. Demaar homepage. Demaar. URL: https://www.demaar.nl/ [accessed 2020-07-10]


Abbreviations

- BMS: Behavioural, Management and Social Sciences
- BRS: Brief Resilience Scale
- CONSORT: Consolidated Standards of Reporting Trials
- EMA: ecological momentary assessment
- HR: human resources
- HRV: heart rate variability
- PSS: Perceived Stress Scale
- RQ: research question
- TIIM: The Incredible Intervention Machine
Perceptions of a Digital Mental Health Platform Among Participants With Depressive Disorder, Anxiety Disorder, and Other Clinically Diagnosed Mental Disorders in Singapore: Usability and Acceptability Study

Ye Sheng Phang¹, BEng; Creighton Heaukulani¹, PhD; Wijaya Martanto¹, MD, PhD; Robert Morris¹,², PhD; Mian Mian Tong³, BSc; Roger Ho³,⁴, MBBS, MD

¹MOH Office for Healthcare Transformation, Singapore, Singapore
²Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
³Department of Psychological Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
⁴Institute for Health Innovation and Technology, National University of Singapore, Singapore, Singapore

Corresponding Author:
Ye Sheng Phang, BEng
MOH Office for Healthcare Transformation
1 Maritime Square
#12-10 Harborfront Centre
Singapore, 099253
Singapore
Phone: 65 6679 3147
Email: yesheng.phang@moht.com.sg

Abstract

Background: The website mindline.sg is a stress management and coping website that can be accessed anonymously in Singapore for free. Although designed to serve individuals who are well or have mild depression and anxiety symptoms, mindline.sg may potentially be used by clinicians as an adjunct therapeutic aid for patients with clinically diagnosed mental disorders.

Objective: This study aims to determine the perceived usability, acceptability, and usefulness of mindline.sg among individuals with diagnosed mental disorders in a clinical setting.

Methods: A cross-sectional study with 173 participants was conducted in the waiting room of a psychiatrist’s office at the National University Hospital in Singapore. Participants waiting for an appointment were given 30 minutes and a simple set of instructions to use three features of mindline.sg. They subsequently answered a set of web-based survey questions via their smartphones, including a 16-item subset of the Post-Study System Usability Questionnaire (PSSUQ) for usability measurement and 5 questions designed to understand the perceived usefulness and acceptability of mindline.sg. Multiple linear regression is used to determine the associated demographic factors with overall PSSUQ score. A chi-square test is performed to investigate associations of psychiatric condition with users’ responses on acceptability and perceived usefulness of mindline.sg. For this study, P<.05 is considered significant.

Results: We observed that the overall (mean 2.86, SD 1.46), system usefulness (mean 2.74, SD 1.46), and information quality (mean 2.98, SD 1.33) subscores of the PSSUQ survey are within a 99% CI of a literature-derived norm, which all have the interpretation of having high perceived usability. However, interface quality (mean 2.98, SD 1.33) scored lower than the literature-derived norm, although it is still better than the neutral score of 4. We find participants with lower than a General Certificate of Education O-Level or N-Level education tend to give a lower usability score as compared to others (β=.49; P=.02). Participants who have not been hospitalized previously due to their condition are also more likely to give a lower PSSUQ score as compared to individuals who have been hospitalized (β=.18; P=.03). The platform mindline.sg is also deemed to be generally useful and acceptable with all the survey questions receiving more than a 60% positive response. We found no association between the type(s) of self-reported psychiatric disorder(s) and the perceived usefulness and acceptability of mindline.sg.

Conclusions: Our results show that mindline.sg is generally perceived as usable and acceptable by individuals with a diagnosed mental disorder in Singapore. The study suggests improving usability among individuals with lower education levels. Particularly promising is the finding that previously hospitalized individuals have significantly higher perceived usability and satisfaction of...
According to the United Nations News, mental disorders affect nearly 1 billion people worldwide [1]. There is also a large treatment gap in mental disorders, which is defined as the difference between the numbers of patients needing and receiving mental health treatment. It was estimated that 76%–85% of the people with severe mental disorders receive no medical treatment in low- and middle-income countries, and that number is around 35%–50% in high-income countries [2]. In Singapore, which is considered a high-income country, 78.6% of individuals met the criteria of needing mental health care but did not receive any treatment or help [3].

The low cost and high accessibility of digital therapeutic tools for mental health have the potential to bridge some of the treatment gaps in Singapore. Indeed, some studies have found evidence to support the use of internet-based mental health self-help tools [4-6].

In June 2020, the Ministry of Health Office for Healthcare Transformation, a subsidiary of the Ministry of Health, launched mindline.sg [7], an anonymous, digital mental health platform. This platform was developed to empower users in Singapore with knowledge, tools, and pathways to self-care as well as resources to help individuals seek out professional help when needed [8]. The website has since rapidly expanded to now include more than 500 curated resources, a self-assessment tool, and an emotionally intelligent artificial intelligence chatbot from Wysa that conducts a range of cognitive behavioral therapy (CBT)–inspired digital therapeutic exercises based on cognitive behavioral therapy. In the 2 years following its launch, mindline.sg received over 485,000 unique visitors.

The platform mindline.sg was not designed to serve individuals with moderate to severe anxiety or depression or those with clinically diagnosed mental disorders. However, it could eventually be expanded to aid health care professionals as an adjunct to therapy for these individuals. To successfully expand mindline.sg to users with diagnosed mental disorders, its usability and acceptability must first be evaluated among this population. Additionally, studies have found that higher acceptability improves uptake and adherence to digital intervention programs [9,10]. This study could also generate insights into product improvement and expansion.

In this study, we aimed to determine the perceived usability and acceptability of mindline.sg among patients with diagnosed mental disorders within a clinical setting. The primary objective of the study was to determine the perceived usability of mindline.sg through the Post-Study System Usability Questionnaire (PSSUQ). The secondary objective was to determine the perceived usefulness and acceptability of mindline.sg through a custom survey.

Methods

Study Design and Recruitment

A cross-sectional study with 173 participants was conducted from April 2021 to January 2022 in the waiting room of a psychiatrist’s office at the Department of Psychological Medicine, National University Hospital, Singapore. Participants waiting for an appointment were given 30 minutes and a simple set of instructions to use three prominent features of mindline.sg: (1) a novel self-assessment and “wellness triaging” questionnaire consisting of a dynamically evolving set of questions from among the Patient Health Questionaire-9 (PHQ-9) and Generalized Anxiety Disorder-7 (GAD-7) surveys, which aims to help users understand their current levels of anxiety or depression and direct them to appropriate content; (2) an emotionally intelligent chatbot from Wysa that conducts a range of cognitive behavioral therapy (CBT)–inspired digital therapeutic tools and can also converse with the user in a free form; and (3) a collection of resources provided by health care ecosystem partners on topics such as mental health literacy, employment support, caregiver support, financial support, fitness tips, and domestic abuse support. Following the 30-minute usage period, the participants answered a set of web-based survey questions, including a reduced version of the PSSUQ with 16 questions designed to measure the perceived usability and satisfaction with the platform and 5 yes/no questions designed to understand the acceptability and perceived usefulness of the website.

The participants were recruited by the Department of Psychological Medicine, National University Hospital in Singapore. Participation was optional, and no remuneration was given. Participants between the ages of 21 and 65 years who were waiting for their psychiatrist appointment at the hospital were invited to the study. For safety reasons, we excluded participants with any form of cancer or major neurological disorder (eg, epilepsy and stroke), heart disease (eg, ischemic heart disease), lung disease (eg, chronic obstructive pulmonary disease), liver disease (eg, liver failure), or kidney disease (eg, kidney failure). All participants were required to provide physical informed consent before the commencement of the study.
Ethics Approval

Ethics approval was granted by the National Healthcare Group Domain Specific Review Board (NHG DSRB 2020/01326) for a study period from January 2021 to January 2022.

Web-Based Survey Design

The web-based survey consisted of 3 sections. In the first section, participant demographic data were collected. This included age, gender, marital status, education level, annual income, employment status, race, and medical history, as shown in Table 1.

Table 1. Demographic data of study participants (N=173).

<table>
<thead>
<tr>
<th>Demographic information of participants</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>66 (38.2)</td>
</tr>
<tr>
<td>Female</td>
<td>107 (61.9)</td>
</tr>
<tr>
<td>Age group (years)</td>
<td></td>
</tr>
<tr>
<td>≤20</td>
<td>1 (0.6)</td>
</tr>
<tr>
<td>21-30</td>
<td>83 (48)</td>
</tr>
<tr>
<td>31-40</td>
<td>40 (23.1)</td>
</tr>
<tr>
<td>41-50</td>
<td>35 (20.2)</td>
</tr>
<tr>
<td>&gt;50</td>
<td>14 (8.1)</td>
</tr>
<tr>
<td>Education level</td>
<td></td>
</tr>
<tr>
<td>Below GCE(^a) O-Level or N-Level</td>
<td>5 (2.9)</td>
</tr>
<tr>
<td>GCE O-Level or N-Level equivalent</td>
<td>19 (11)</td>
</tr>
<tr>
<td>Diploma, A-Level, or equivalent</td>
<td>66 (38.2)</td>
</tr>
<tr>
<td>Undergraduate degree and above</td>
<td>79 (45.7)</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>115 (66.5)</td>
</tr>
<tr>
<td>Married</td>
<td>58 (33.5)</td>
</tr>
<tr>
<td>Annual income (SGD)(^b)</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>49 (28.3)</td>
</tr>
<tr>
<td>&lt;30,000</td>
<td>46 (26.6)</td>
</tr>
<tr>
<td>30,000-60,000</td>
<td>49 (28.3)</td>
</tr>
<tr>
<td>60,001-100,000</td>
<td>15 (8.7)</td>
</tr>
<tr>
<td>&gt;100,000</td>
<td>14 (8.1)</td>
</tr>
<tr>
<td>Race</td>
<td></td>
</tr>
<tr>
<td>Chinese</td>
<td>119 (68.8)</td>
</tr>
<tr>
<td>Malay</td>
<td>24 (13.9)</td>
</tr>
<tr>
<td>Indian</td>
<td>14 (8.1)</td>
</tr>
<tr>
<td>Other</td>
<td>16 (9.3)</td>
</tr>
</tbody>
</table>

\(^a\)GCE: General Certificate of Education.

\(^b\)A currency exchange rate of SGD $1=US $0.74 is applicable.

The second section was modelled after the reduced version of the PSSUQ, which is a widely deployed usability quantification survey [11]. The PSSUQ consists of 16 questions that are scored on a 7-point scale (from 1 as “strongly agree” to 7 as “strongly disagree”; Multimedia Appendix 1). The scores determine an overall satisfaction scale, computed as the average score across all 16 items (and so takes a value between 16 and 112) and 3 subscales of system usefulness, information quality, and interface quality, taking values in the ranges of 6-42, 6-42, and 4-24, respectively, each computed as the average score across various subsets of the items. For all scales, lower scores indicate better usability. The PSSUQ questionnaire has shown a satisfactory level of reliability, sensitivity, and validity [12,13].

The third section of the survey consisted of the following 5 yes/no questions constructed by the study team to measure the
acceptability and perceived usefulness of the major features of mindline.sg:

1. Did you like taking the “I need help to manage my emotions” questionnaires?
2. Did you find the resources that are listed useful?
3. Did you like talking with the Wysa chatbot (the penguin)?
4. Did you find any of the exercises recommended by mindline.sg or Wysa (the emotionally intelligent chatbot) useful?
5. Would you recommend mindline.sg to a friend?

These questions have not been tested for reliability, sensitivity, and validity.

**Statistical Analysis**

We used multiple linear regression analyses to discover factors that are associated with the overall PSSUQ scale, such as age, marital status, annual income, education level, and medical history. The empirical distribution of the overall PSSUQ scale was not well modelled by a Gaussian distribution (Shapiro-Wilks test: \( P < .001 \)), so a power transformation was applied to the PSSUQ scales before training.

To investigate any relationship between the self-reported psychiatric conditions and the responses to the 5 yes/no survey questions measuring the acceptability and perceived usefulness of the platform, we performed chi-square tests based on self-reported psychiatric conditions. Three population comparisons were performed: (1) between participants with and without depressive disorder, (2) between participants with and without anxiety disorder, and (3) between participants diagnosed with any other psychiatric conditions and those diagnosed with either depressive or anxiety disorders. We noted that 10 participants indicated they either did not have any psychiatric disorder or did not know their diagnosis; the responses of these participants were excluded from this analysis.

All statistical tests were 2-tailed. We reported factors at \( P < .05 \) as significant, and corrections for multiple comparisons were not used.

**Results**

**Demographic Data**

The demographic data collected, summarized in Table 1, revealed the majority of the participants were female (107/173, 61.9%), between the ages of 21 and 30 years (83/173, 48%), had an education level of “undergraduate degree and above” (79/173, 45.7%), and were single (115/173, 66.5%).

**Medical History**

The medical history data are summarized in Table 2. The majority of the participants (88/173, 50.8%) self-reported being diagnosed with depressive disorder, were most likely to be on either medication (84/173, 48.6%) or both medication and psychotherapy or counselling (71/173, 41%), and had not been hospitalized due to their psychiatric condition (107/173, 61.9%).
Table 2. Data on the medical history of the participants.

<table>
<thead>
<tr>
<th>Medical history</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do you suffer from any psychiatric condition?</td>
<td></td>
</tr>
<tr>
<td>Depressive disorder</td>
<td>88 (50.8)</td>
</tr>
<tr>
<td>Anxiety disorder</td>
<td>38 (22)</td>
</tr>
<tr>
<td>Both anxiety and depressive disorder</td>
<td>14 (8.1)</td>
</tr>
<tr>
<td>Others (e.g., bipolar disorder, attention deficit hyperactivity disorder, adjustment disorder, schizophrenia, borderline personality disorder, and alcoholism)</td>
<td>23 (13.3)</td>
</tr>
<tr>
<td>Unknown or undiagnosed</td>
<td>10 (5.8)</td>
</tr>
<tr>
<td>How long (in years) have you suffered from this psychiatric condition(s)?</td>
<td></td>
</tr>
<tr>
<td>&lt;1</td>
<td>29 (17.8)</td>
</tr>
<tr>
<td>≥1 to &lt;5</td>
<td>44 (27)</td>
</tr>
<tr>
<td>≥5 to &lt;10</td>
<td>40 (24.6)</td>
</tr>
<tr>
<td>≥10 to &lt;15</td>
<td>27 (16.6)</td>
</tr>
<tr>
<td>≥15</td>
<td>23 (14.1)</td>
</tr>
<tr>
<td>What kind of treatment are you receiving for this psychiatric condition(s)?</td>
<td></td>
</tr>
<tr>
<td>Medication</td>
<td>84 (48.6)</td>
</tr>
<tr>
<td>Psychotherapy or counseling</td>
<td>2 (1.2)</td>
</tr>
<tr>
<td>Both medication and psychotherapy or counselling</td>
<td>71 (41)</td>
</tr>
<tr>
<td>Others</td>
<td>6 (3.5)</td>
</tr>
<tr>
<td>Not under any treatment</td>
<td>10 (5.8)</td>
</tr>
<tr>
<td>Have you been hospitalized due to this psychiatric condition(s)?</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>66 (38.2)</td>
</tr>
<tr>
<td>No</td>
<td>107 (61.9)</td>
</tr>
<tr>
<td>Have you been on depression medication?</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>133 (77.3)</td>
</tr>
<tr>
<td>No</td>
<td>40 (23.1)</td>
</tr>
<tr>
<td>Do you suffer from any other chronic medical conditions?</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>48 (27.8)</td>
</tr>
<tr>
<td>No</td>
<td>125 (72.3)</td>
</tr>
</tbody>
</table>

*aUnknown or undiagnosed participants are not required to answer this question.

PSSUQ Results

In Table 3, we report the mean and SD values of the PSSUQ overall score and subscores. A meta-analysis of 5 years of usability studies (which were predominantly on speech recognition systems, though the meta-analysis showed a good ability to generalize) provided the means and 99% CIs of analyzed PSSUQ scores (Table 3) [11]. We will henceforth refer to the literature-derived mean and 99% CIs as PSSUQ norms. A lower score indicates better usability.

In Table 4, we report the parameter estimates in a multiple linear regression model (a Box-Cox transformation) of the PSSUQ overall satisfaction scale onto the participant demographic and medical history data.

Table 3. The overall Post-Study System Usability Questionnaire (PSSUQ) scores and subscores as well as the norms.

<table>
<thead>
<tr>
<th>Questions</th>
<th>PSSUQ score, mean (SD)</th>
<th>PSSUQ norms, mean (99% CI) [10]</th>
</tr>
</thead>
<tbody>
<tr>
<td>System usefulness</td>
<td>2.74 (1.46)</td>
<td>2.80 (2.57-3.02)</td>
</tr>
<tr>
<td>Information quality</td>
<td>2.98 (1.33)</td>
<td>3.02 (2.79-3.24)</td>
</tr>
<tr>
<td>Interface quality</td>
<td>2.82 (1.59)</td>
<td>2.49 (2.28-2.71)</td>
</tr>
<tr>
<td>Overall</td>
<td>2.86 (1.46)</td>
<td>2.82 (2.62-3.02)</td>
</tr>
</tbody>
</table>
Table 4. Parameter estimates in multiple linear regression (a power transformation) of the overall satisfaction with the Post-Study System Usability Questionnaire (PSSUQ) scale onto the participant demographic and medical history data. Italicized $P$ values are significant.

<table>
<thead>
<tr>
<th>Factors associated with the overall satisfaction PSSUQ score</th>
<th>$\beta$</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>-.00</td>
<td>.76</td>
</tr>
<tr>
<td>Gender (male vs female)</td>
<td>-.00</td>
<td>&gt;.99</td>
</tr>
<tr>
<td>Marital status (married vs single)</td>
<td>.155</td>
<td>.13</td>
</tr>
<tr>
<td><strong>Annual income (vs no income; SGD)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;30,000</td>
<td>.175</td>
<td>.09</td>
</tr>
<tr>
<td>30,000-60,000</td>
<td>.201</td>
<td>.06</td>
</tr>
<tr>
<td>60,001-100,000</td>
<td>.205</td>
<td>.23</td>
</tr>
<tr>
<td>&gt;100,000</td>
<td>.132</td>
<td>.45</td>
</tr>
<tr>
<td><strong>Education level (vs university degree and above)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Below GCE&lt;sup&gt;b&lt;/sup&gt; O-Level or N-Level certification or equivalent</td>
<td>.491</td>
<td>.02</td>
</tr>
<tr>
<td>GCE O-Level or N-Level certification or equivalent</td>
<td>.044</td>
<td>.75</td>
</tr>
<tr>
<td>Diploma, A-Level, or equivalent</td>
<td>.053</td>
<td>.58</td>
</tr>
<tr>
<td><strong>Medical condition (vs others)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depressive disorder</td>
<td>.137</td>
<td>.13</td>
</tr>
<tr>
<td>Anxiety disorder</td>
<td>-.063</td>
<td>.51</td>
</tr>
<tr>
<td>Have not been previously hospitalized for a psychiatric condition (vs having been previously hospitalized)</td>
<td>.177</td>
<td>.03</td>
</tr>
</tbody>
</table>

<sup>a</sup>A currency exchange rate of SGD $1=US $0.74 is applicable.

<sup>b</sup>GCE: General Certificate of Education.

Acceptability and Perceived Usefulness of “mindline.sg”

The last section of the survey consists of five yes/no questions designed to measure the acceptability and perceived usefulness of mindline.sg. The overall responses were largely positive with all the 5 questions receiving more than a 60% positive response. The question “Did you find the resources that are listed useful?” received the highest percentage of positive responses (86.7% of the users). The question “Did you find any of the exercises recommended by mindline.sg or Wysa useful?” received the lowest percentage of positive response (60%). Table 5 shows the full results of the survey. We found no significant differences between the responses of any subgroups.
Table 5. An analysis comparing the distributions of the responses to the survey questions measuring the usefulness and acceptability of the platform between participants segmented into three groups: (1) participants with depressive disorder (DD) versus participants without DD; (2) participants with anxiety disorder (AD) versus participants without AD; and (3) participants with other psychiatric conditions, excluding DD and AD, versus participants with DD and AD. The P value of chi-square tests comparing the various subpopulation distributions is reported in the final column. There are no differences in survey responses between the diagnoses at significance level (P=.05).

<table>
<thead>
<tr>
<th>Items</th>
<th>Total (N=163), n (%)</th>
<th>DD vs no DD</th>
<th>AD vs no AD</th>
<th>Others vs DD and AD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DD (n=102), n (%)</td>
<td>No DD (n=61), n (%)</td>
<td>Chi-square (df)</td>
<td>P value</td>
</tr>
<tr>
<td>Did you like taking the “I need help to manage my emotions” questionnaires?</td>
<td>No</td>
<td>31 (19.7)</td>
<td>22 (21.6)</td>
<td>9 (14.8)</td>
</tr>
<tr>
<td>Yes</td>
<td>132 (80.4)</td>
<td>80 (78.4)</td>
<td>52 (85.2)</td>
<td>—</td>
</tr>
<tr>
<td>Did you find the resources that are listed useful?</td>
<td>No</td>
<td>23 (13.3)</td>
<td>14 (13.7)</td>
<td>9 (14.6)</td>
</tr>
<tr>
<td>Yes</td>
<td>140 (86.7)</td>
<td>88 (86.3)</td>
<td>52 (85.2)</td>
<td>—</td>
</tr>
<tr>
<td>Did you like talking with the Wysa chatbot (the penguin)?</td>
<td>No</td>
<td>53 (31.8)</td>
<td>37 (36.3)</td>
<td>16 (26.2)</td>
</tr>
<tr>
<td>Yes</td>
<td>110 (68.2)</td>
<td>65 (63.7)</td>
<td>45 (73.8)</td>
<td>—</td>
</tr>
<tr>
<td>Did you find any of the exercises recommended by mindline.sg or Wysa useful?</td>
<td>No</td>
<td>63 (39.9)</td>
<td>45 (44.1)</td>
<td>18 (29.5)</td>
</tr>
<tr>
<td>Yes</td>
<td>100 (60.1)</td>
<td>57 (55.9)</td>
<td>43 (70.5)</td>
<td>—</td>
</tr>
<tr>
<td>Would you recommend mindline.sg to a friend?</td>
<td>No</td>
<td>40 (24.3)</td>
<td>29 (28.4)</td>
<td>11 (18)</td>
</tr>
<tr>
<td>Yes</td>
<td>123 (75.7)</td>
<td>73 (71.6)</td>
<td>50 (81)</td>
<td>—</td>
</tr>
</tbody>
</table>

aNot applicable.

Discussion

Usability Findings

Comparing the PSSUQ overall score (describing the perceived usability of mindline.sg among the survey respondents) to the literature-derived norms (Table 3), we found that the system usefulness (mean 2.74, SD 1.46), information quality (mean 2.98, SD 1.33), and the platform overall score (mean 2.86, SD 1.46) were perceived as “good” and were comparable to most other digital apps (within a 99% CI of the literature-derived norm). Although the interface quality score (mean 2.98, SD 1.33) is lower in this regard than most other digital apps, it is also perceived as “good” because it is above the neutral score of 4 on the PSSUQ scale.

Based on the results of the multiple linear regression with the overall PSSUQ scores, we found that education level is the factor with the highest association with the PSSUQ score (the largest magnitude coefficient is reported in Table 4). In particular, participants with an education lower than an O-Level or N-Level General Certificate of Education tend to give mindline.sg a lower usability score (β=0.49; P=.02). Because the majority of the resources on mindline.sg are text-based and require a certain level of English literacy to use, we find the results in line with our expectations. However, these findings are in contrast with earlier studies, which found a weak to no correlation between education level and usability of connected medical devices and internet-based CBT (iCBT) platforms [10,13].

In addition, participants who have not been previously hospitalized due to their psychiatric condition are likely to give a poorer overall PSSUQ score compared to participants who...
have been hospitalized ($\beta=0.177; P=.03$). At the point of writing, we could not find any prior study to explain this finding.

**Acceptability and Perceived Usefulness Findings**

We found that mindline.sg is generally acceptable to participants with self-reported mental disorders, with all 5 questions having more than 60% positive response. These results are consistent with the findings from a meta-analysis that found iCBT platforms were acceptable and effective for patients with depression and anxiety disorders [14].

When comparing patients' responses on mindline.sg, acceptability and perceived usefulness between the three different mental condition subgroups (as illustrated in Table 5), we found no significant differences between responses by type of psychiatric disorder.

**Limitations**

The participants were given only 30 minutes to use the three features of mindline.sg before they were asked to complete the survey, but some of the therapeutic exercises provided by the Wysa chatbot on the website could take around 20-30 minutes to complete. This could explain why the questions on “Did you find any of the exercises recommended by mindline.sg or Wysa useful?” received the least amount of positive response (60%) compared to the other acceptability questions, as the participants may not have had enough time to fully explore these exercises. Other resources are usually completed in a shorter amount of time, as they take the form of articles and videos that generally take less than 10 minutes to consume. The short usage period of mindline.sg in this study may affect the generalizability of this study.

Additionally, the medical history collected in this survey is self-reported and has not been independently verified with clinical records (this was not put forward to the ethics committee to protect patient confidentiality and preserve the anonymity feature of the mindline.sg platform). The data reported in our study are also from participants who had an appointment in the Department of Psychological Medicine in National University Hospital and were not randomly selected, which could result in some form of selection bias. Given the limitations mentioned above, any generalization from this study should be evaluated with caution.

Although the PSSUQ norms were used as a basis for comparison with our collected results, it is important to note that the norms were established by products from a variety of sources (which were predominantly speech recognition systems) and at different stages of development [12].

Although we compare our results of acceptability to a meta-analysis, it is also important to note that many of the studies in the meta-analysis use adherence and patient satisfaction in a longer-term treatment program as a proxy for acceptability [13]. Since our participants only use mindline.sg for around 30 minutes, the answer to the 5 questions is instead used as a proxy to acceptability.

Lastly, as the evaluation of usability in mobile health varies substantially [15], it presents a challenge for us to compare our findings to previously published usability results. Although the PSSUQ questionnaire has shown a satisfactory level of reliability, sensitivity, and validity [11,12], the 5 yes/no questions that were constructed by the study team as a proxy for acceptability have not been tested for validity and reliability. The nature of the yes/no questions could also limit the range of responses as compared to a Likert-type scale.

**Conclusions**

Despite the limitations mentioned above, this study shows that mindline.sg could be a viable self-help tool for individuals with diagnosed mental health conditions due to its well-rated usability and acceptability. Furthermore, the accessibility of a free, anonymous, and web-based tool like mindline.sg allows people with diagnosed mental conditions to access these services at any time and from the comfort and privacy of their homes. However, the clinical effectiveness of mindline.sg as a mental health resource for people diagnosed with mental conditions has not yet been validated and might be an important focus for future studies.

**Authors’ Contributions**

YSP, WM, RM, MMT, and RH designed the study. RH and MMT provided human resources and prepared the venue for data collection. YSP and CH developed the outline and contributed to analyses, interpreted results, and wrote the first and final draft of the manuscript. All authors read and approved the final manuscript.

**Conflicts of Interest**

The website mindline.sg is developed and maintained by the MOH Office for Healthcare Transformation. YSP, CH, WM, and RM were employees of MOH Office for Healthcare Transformation during the period of study.

**Multimedia Appendix 1**

Poststudy usability questionnaire used for this study.
[DOCX File, 21 KB - humanfactors_v10i1e42167_app1.docx ]

**References**

Abbreviations

CBT: cognitive behavioral therapy
GAD-7: Generalized Anxiety Disorder-7
iCBT: internet-based cognitive behavioral therapy
MOH: Ministry of Health
PHQ-9: Patient Health Questionnaire-9
PSSUQ: Post-Study System Usability Questionnaire

Edited by A Kushniruk; submitted 25.08.22; peer-reviewed by YJ Bao, L. Sequeira; comments to author 08.11.22; revised version received 24.12.22; accepted 11.02.23; published 29.03.23.

Please cite as:

Phang YS, Heakaluddin C, Martanto W, Morris R, Tong MM, Ho R
Perceptions of a Digital Mental Health Platform Among Participants With Depressive Disorder, Anxiety Disorder, and Other Clinically Diagnosed Mental Disorders in Singapore: Usability and Acceptability Study
JMIR Hum Factors 2023;10:e42167
URL: https://humanfactors.jmir.org/2023/1/e42167
PMID: 36989920
Evaluating User Experience With a Chatbot Designed as a Public Health Response to the COVID-19 Pandemic in Brazil: Mixed Methods Study

Bruno Azevedo Chagas, PhD; Adriana Silvina Pagano, PhD; Raquel Oliveira Prates, PhD; Elisa Cordeiro Praes, DVM; Kicila Ferreguetti, PhD; Helena Vaz; Zilma Silveira Nogueira Reis, PhD; Leonardo Bonisson Ribeiro, BSc; Antonio Luiz Pinho Ribeiro, PhD; Thais Marques Pedroso, PhD; Alline Beleigoli, DVM; Clara Rodrigues Alves Oliveira, PhD; Milena Soriano Marcolino, PhD

1Computer Science Department, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
2Arts Faculty, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
3Telehealth Center, University Hospital, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
4Telehealth Network of Minas Gerais, Belo Horizonte, Brazil
5Department of Internal Medicine, Medical School, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
6Flinders Digital Health Centre, Flinders University, Adelaide, Australia
7Caring Futures Institute, Flinders University, Adelaide, Australia

Corresponding Author:
Milena Soriano Marcolino, PhD
Telehealth Center
University Hospital
Universidade Federal de Minas Gerais
Avenida Professor Alfredo Balena
110 1o andar Sala 107 Ala Sul
Belo Horizonte, 30130-100
Brazil
Phone: 55 31 3307 9201
Email: milenamarc@gmail.com

Abstract

Background: The potential of chatbots for screening and monitoring COVID-19 was envisioned since the outbreak of the disease. Chatbots can help disseminate up-to-date and trustworthy information, promote healthy social behavior, and support the provision of health care services safely and at scale. In this scenario and in view of its far-reaching postpandemic impact, it is important to evaluate user experience with this kind of application.

Objective: We aimed to evaluate the quality of user experience with a COVID-19 chatbot designed by a large telehealth service in Brazil, focusing on the usability of real users and the exploration of strengths and shortcomings of the chatbot, as revealed in reports by participants in simulated scenarios.

Methods: We examined a chatbot developed by a multidisciplinary team and used it as a component within the workflow of a local public health care service. The chatbot had 2 core functionalities: assisting web-based screening of COVID-19 symptom severity and providing evidence-based information to the population. From October 2020 to January 2021, we conducted a mixed methods approach and performed a 2-fold evaluation of user experience with our chatbot by following 2 methods: a posttask usability Likert-scale survey presented to all users after concluding their interaction with the bot and an interview with volunteer participants who engaged in a simulated interaction with the bot guided by the interviewer.

Results: Usability assessment with 63 users revealed very good scores for chatbot usefulness (4.57), likelihood of being recommended (4.48), ease of use (4.44), and user satisfaction (4.38). Interviews with 15 volunteers provided insights into the strengths and shortcomings of our bot. Comments on the positive aspects and problems reported by users were analyzed in terms of recurrent themes. We identified 6 positive aspects and 15 issues organized in 2 categories: usability of the chatbot and health support offered by it, the former referring to usability of the chatbot and how users can interact with it and the latter referring to the chatbot’s goal in supporting people during the pandemic through the screening process and education to users through...
informative content. We found 6 themes accounting for what people liked most about our chatbot and why they found it useful—3 themes pertaining to the usability domain and 3 themes regarding health support. Our findings also identified 15 types of problems producing a negative impact on users—10 of them related to the usability of the chatbot and 5 related to the health support it provides.

Conclusions: Our results indicate that users had an overall positive experience with the chatbot and found the health support relevant. Nonetheless, qualitative evaluation of the chatbot indicated challenges and directions to be pursued in improving not only our COVID-19 chatbot but also health chatbots in general.

(JMIR Hum Factors 2023;10:e43135) doi:10.2196/43135

KEYWORDS
user experience; chatbots; telehealth; COVID-19; human-computer interaction; HCI; empirical studies in human-computer interaction; empirical studies in HCI; health care information systems

Introduction

The burden on health systems during the COVID-19 pandemic reached unprecedented levels in both high- and low-income countries globally. The increase in demand for the provision of care through the several COVID-19 pandemic waves required global public health responses and challenged health care systems’ capacity as well as health units’ resilience [1]. Concomitantly, there was a sudden unprecedented demand for information and a widespread amount of unreliable and fake information—an “infodemic” [2]—putting lives at risk by prompting the population to try unproven medications in the hope of preventing the disease or finding a “cure” [3]. In this context, telehealth and digital health solutions, including chatbots, emerged as a quick and viable response, acting as a symptom checker in digital triage approaches [1,4,5].

Chatbots are conversational agents that interact with people using a text-based interface or spoken natural language [6]. They are usually deployed through website widgets or instant messaging apps and have been increasingly adopted in several different fields such as finance, commerce, marketing, and fitness [7]. They have only recently started to expand into health care [8]. Their method of communication makes it suitable for a variety of target populations; various health conditions; and a broad range of purposes such as patient triage, clinical decision support, and self-management [9-12].

The potential of chatbots for screening and monitoring COVID-19 was envisioned since the disease outbreak as a strategy not only to disseminate up-to-date and trustworthy information but also to promote healthy social behavior and to support the provision of health care services safely and at scale [13]. For the purpose of pandemic management, chatbots might teach people about social distancing and other prevention measures; clarify doubts about symptoms, treatments, and vaccines; and help screen patients remotely, avoiding unnecessary visits to health care centers that could implicate crowding and taking up valuable time of health care professionals [14].

In this scenario and in view of its far-reaching postpandemic impact, it is critically important to evaluate user experience with this type of technology. Despite the World Health Organization (WHO) recommendation regarding the assessment of user interaction for the adoption of digital technologies in health care, evidence on chatbot assessment in the context of the COVID-19 pandemic and other conditions is still scarce [4,5,15]. This is of utmost importance not only as a way to assess and enhance users’ experiences but also to improve the technology itself, so that it can fulfill its ultimate goal of promoting public health and saving lives even during a scenario of uncertainties from the lack of evidence and ethical risks. In addition, assessment can provide insights for the development of chatbots for other conditions. The better the quality of user experience, the greater the chances of adoption and benefits for most users.

Therefore, this paper sought to evaluate the quality of user interaction with a chatbot developed to respond to the COVID-19 pandemic by a large telehealth service in Brazil to assess users’ overall experiences, including strengths and shortcomings, as reported by participants.

Methods

Chatbot Development and Implementation

The planning and development of our COVID-19 chatbot were described in detail previously [16,17]. The bot was developed in March 2020 at the beginning of the first wave of the COVID-19 pandemic in Brazil to provide 2 core functionalities. The first was assisting web-based screening of COVID-19 symptom severity based on a decision tree that considered available evidence and recommendations from the Brazilian Ministry of Health [18] and the WHO [19]. This functionality was meant to (1) advise the population whether and when to seek care, with people with no warning signs advised to stay home; and (2) queue patients for teleconsultation, prioritizing those with warning sign severity and comorbidities [20]. Figure 1 shows a flowchart of the stages the user traverses guided by the chatbot questions. Colors are used to screen cases: (1) red (user advised to search for immediate, emergency care); (2) orange (user advised to search for urgent care at the hospital); (3) yellow (user advised to search for care in reference centers); and (4) green (user advised to stay at home unless new warning signs appear).

The second functionality aimed to supply evidence-based information to the population at a time of uncertainty, misinformation, and widespread dissemination of fake news. Misleading information can be created and used unintentionally or intentionally to cause harm (misinformation vs disinformation vs malinformation) [21]. However, there is misleading...
information from the lack of consistent evidence regarding many aspects of this recent disease, which demanded continuous revision in the scientific basis of the chatbot. This was provided as question and answer (Q&A) based on frequently asked questions in the database at the Telehealth Center of the University Hospital at Universidade Federal de Minas Gerais [22]. The questions were initially grouped into 11 topics—general information, transmission, symptoms, advice for suspected cases, treatment, home care, hygiene, lifestyle, mask use, pregnancy, and pet care—and later expanded to include diagnosis. A group of health care professionals at the Telehealth Center selected 85 Q&A pairs based on the best available evidence and following the Brazilian Ministry of Health [18] and WHO [19] recommendations.

Our chatbot, having a female identity and the name Ana, was developed using BLiP [23]—a proprietary software platform—as a service for the development of conversational agents. The chatbot was available via different channels, namely as an app on WhatsApp (Meta Platforms Inc); as a webchat on the web sites maintained by the Telehealth Center [24] (Figure 2), the city of Teófilo Otoni [25], and the University of São João del Rei in Divinópolis [26]; and as an “embedded” app hosted by Divinópolis municipal health department [20]. A version of the chatbot with a male identity and the name Pedro was also made available on the website maintained by Universidade Federal de Minas Gerais for students and personnel to queue for teleconsultations and have access to frequently asked questions. For the purposes of our study, we focused solely on the chatbot Ana.

Figure 1. Decision tree for screening suspect cases of COVID-19.

Figure 2. Screenshot of the Telehealth Center website showing our chatbot Ana as a widget at the bottom right of the page.
Study Design

A mixed methods approach was used, and user experience with the chatbot was evaluated through (1) a posttask usability survey administered to a sample of users who resorted to the bot for symptom checking to gather participants’ impressions immediately after concluding their interaction with the bot and (2) an interview with volunteer participants who engaged in simulated interaction with the bot guided by the interviewer. We performed a convergent parallel mixed methods design [27], in which data were collected and analyzed separately, and the results were presented side by side and then related at the end. Both studies address the same macro–research question regarding user experience with the chatbot. The quantitative study is meant to indicate a broad trend, whereas the qualitative study is meant to provide deeper insight into the user experience.

Ethics Approval

The study protocol was approved by the Brazilian National Commission for Research Ethics (CAAE 35953620.9.0000.5149). Individual informed consent was obtained for all the participants.

Usability Survey

A brief usability survey was used to assess users’ overall impressions after they had concluded using the chatbot. The survey was intended to evaluate chatbot usability at scale and was administered to all users after concluding their interaction with the chatbot. As they were symptomatic users who were actually quite concerned about their health condition and were not very willing to spend time answering a questionnaire, we opted to use a small set of 4 questions drawing on the classic criteria for usability assessment [28,29]. The questions inquired on 4 usability aspects, namely ease of use, usefulness, satisfaction, and likelihood of recommending the bot to other users. Answers were collected using a 5-point Likert scale, ranging from 1 (worst score) to 5 (best score), representing the strength with which the respondent agreed or disagreed with each question. All users were invited to reply to the survey, but replying was optional, and users could comply and accept the invitation or conclude their chatbot session without answering our survey. From October 2020 to January 2021, 622 complete interactions with the chatbot were recorded. In total, 63 out of 622 users agreed to fill in our usability survey (response rate of 10.1%). Table 1 shows the sociodemographic data of respondents and nonrespondents.

Table 1. User profile of recorded interactions.a.

<table>
<thead>
<tr>
<th>Users</th>
<th>Age (years), mean (SD)</th>
<th>Women (n=380), n (%)</th>
<th>Men (n=237), n (%)</th>
<th>Not declared (n=5), n (%)</th>
<th>Total (n=622), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Respondents</td>
<td>36.1 (14.4)</td>
<td>46 (12.1)</td>
<td>17 (7.2)</td>
<td>0 (0)</td>
<td>63 (10.1)</td>
</tr>
<tr>
<td>Nonrespondents</td>
<td>34.5 (14.1)</td>
<td>334 (87.9)</td>
<td>220 (92.8)</td>
<td>5 (100)</td>
<td>559 (89.9)</td>
</tr>
<tr>
<td>Total</td>
<td>34.7 (14.1)</td>
<td>380 (100)</td>
<td>237 (100)</td>
<td>5 (100)</td>
<td>622 (100)</td>
</tr>
</tbody>
</table>

aInformation was recorded during interaction as informed by users.

The respondents had a mean age of 36.1 (SD 14.4) years and were predominantly women (46 out of 63 respondents, 73%).

Descriptive statistics assessed the characteristics of the users and responses to the usability questions. To summarize the quantitative variables, we used averages, SDs, medians, minimum and maximum, or IQRs depending on the data distribution. Qualitative variables were presented as absolute values and percentages. Box plots enhanced the visualization of grades assigned by users on each criterion for assessment.

Qualitative Assessment: Users’ Interviews and Analysis

To tap users’ assessment of the chatbot interface and core functionalities (screening and educational session), we conducted a remote teleconference session with 15 invited volunteer asymptomatic participants having different age, sex, and occupation profiles recruited by the research team. Each participant received a scenario describing a situation that would prompt their interaction with the chatbot. The researcher observed and recorded their interaction. The session was followed by a semistructured interview to gather insights on their experience with the chatbot and their perceptions of the strengths and shortcomings of the bot as reported by them.

The evaluation was conducted through a teleconference system and took place between November 2020 and January 2021 as the second wave of the pandemic started in Brazil. The interviews were transcribed, and a thematic analysis was performed [30].

Among the 15 participants, 53% (n=8) were female, with ages ranging from 18 to 62 (mean 38.1, SD 15.7; median 37; minimum=18, Q1=25, Q3=51, maximum=62) years, and 73% (n=11) had a higher education degree. Out of the participants, 33% (n=5) were engaged in teaching or research at the university, 27% (n=4) were students and 40% (n=6) were regular or self-employed workers. With regard to the device used to interact with the chatbot, 80% (n=12) used a desktop or laptop computer, whereas 20% (n=3) used a smartphone. The participants’ data are detailed in Table 2.

In the evaluation session, the participants received a scenario describing a situation that would prompt their interaction with the chatbot. A set of 10 different scenarios were prepared to cover different chatbot interactive paths in the screening functionality, from severe to light symptoms, with and without comorbidities (Multimedia Appendix 1). Participants were designated to scenarios according to their actual profiles to make the interaction as realistic as possible. Sample scenarios included an adult woman in her 30s being assigned a scenario of a pregnant woman, a participant in their 60s being assigned a scenario of a person with some comorbidity, among others. Similarly, each scenario included 3 topics to assess the educational functionality of the chatbot, 2 of them being...
preassigned topics, and a third one free for the participant to choose. Most sessions lasted between 30 minutes and 1 hour. During the sessions, the participants interacted with the chatbot while the researcher observed and recorded their interactions. Afterward, they were interviewed about their experience with the chatbot (the interview script used is available in Multimedia Appendix 2). The evaluation was conducted through a teleconference system chosen by the participant, in an individual session and in Portuguese (participants’ mother tongue), which took place between November 2020 and January 2021.

Table 2. Profile of participants taking part in the semistructured interview.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Age (years)</th>
<th>Sex</th>
<th>Education</th>
<th>Occupation</th>
<th>Device used</th>
</tr>
</thead>
<tbody>
<tr>
<td>P01</td>
<td>37</td>
<td>Male</td>
<td>Graduate degree</td>
<td>University lecturer</td>
<td>Desktop or laptop computer</td>
</tr>
<tr>
<td>P02</td>
<td>48</td>
<td>Female</td>
<td>Graduate degree</td>
<td>University lecturer</td>
<td>Desktop or laptop computer</td>
</tr>
<tr>
<td>P03</td>
<td>25</td>
<td>Male</td>
<td>Graduate degree</td>
<td>Attorney</td>
<td>Desktop or laptop computer</td>
</tr>
<tr>
<td>P04</td>
<td>40</td>
<td>Male</td>
<td>Graduate degree</td>
<td>IT or user experience designer</td>
<td>Smartphone</td>
</tr>
<tr>
<td>P05</td>
<td>25</td>
<td>Female</td>
<td>Bachelor’s degree in linguistics</td>
<td>Student pursuing a master’s degree</td>
<td>Desktop or laptop computer</td>
</tr>
<tr>
<td>P06</td>
<td>58</td>
<td>Female</td>
<td>Bachelor’s degree in nutrition science</td>
<td>Credentialed dietitian and undergraduate student in psychology</td>
<td>Smartphone</td>
</tr>
<tr>
<td>P07</td>
<td>27</td>
<td>Female</td>
<td>Bachelor’s degree in veterinary studies</td>
<td>Undergraduate student in linguistics</td>
<td>Desktop or laptop computer</td>
</tr>
<tr>
<td>P08</td>
<td>52</td>
<td>Female</td>
<td>Graduate degree</td>
<td>Lecturer</td>
<td>Desktop or laptop computer</td>
</tr>
<tr>
<td>P09</td>
<td>33</td>
<td>Male</td>
<td>Graduate degree</td>
<td>Sociologist</td>
<td>Desktop or laptop computer</td>
</tr>
<tr>
<td>P10</td>
<td>50</td>
<td>Female</td>
<td>Graduate degree</td>
<td>Psychologist</td>
<td>Desktop or laptop computer</td>
</tr>
<tr>
<td>P11</td>
<td>20</td>
<td>Female</td>
<td>High school degree</td>
<td>Undergraduate student in psychology</td>
<td>Desktop or laptop computer</td>
</tr>
<tr>
<td>P12</td>
<td>18</td>
<td>Female</td>
<td>High school degree</td>
<td>Student</td>
<td>Desktop or laptop computer</td>
</tr>
<tr>
<td>P13</td>
<td>59</td>
<td>Male</td>
<td>Bachelor’s degree in computer science</td>
<td>IT analyst</td>
<td>Desktop or laptop computer</td>
</tr>
<tr>
<td>P14</td>
<td>18</td>
<td>Male</td>
<td>High school degree</td>
<td>Undergraduate student</td>
<td>Desktop or laptop computer</td>
</tr>
<tr>
<td>P15</td>
<td>62</td>
<td>Male</td>
<td>High school degree</td>
<td>Insurance broker</td>
<td>Smartphone</td>
</tr>
</tbody>
</table>

*a* Master’s or doctoral degree.

The interviews were recorded and included screen recordings of participants’ interactions with the chatbot. The interviews were transcribed by the research team. Thematic analysis [30] of the interview transcripts was carried out to find recurrent themes in participants’ interviews that could be matched to the research questions guiding our study as follows:

1. What are the strengths and shortcomings of our bot as perceived by users?
2. What particular insights can be drawn from our study to inform prospective chatbot design?

Our thematic analysis was conducted in an inductive way, that is, a bottom-up approach, where the analysis is not driven by a preexisting framework or theory, but the researchers search for codes and themes in a data-driven way [30]. This approach is applicable for qualitative analysis of interview data [31] and is more suitable for broad rather than specific research questions, as was our case [30].

We applied triangulation as a typical strategy to improve the quality and reliability of our qualitative results [32]. In particular, the data were analyzed by multiple researchers (investigator triangulation [33,34]) and the outcome of their analysis was discussed until consensus was reached. The transcripts were coded by 2 senior and 2 junior researchers, with a set of at least 5 transcripts being assigned to each one for analysis and coding. Thus, every interview was analyzed by at least 2 different researchers. Interviews were recorded and analyzed in the qualitative data analysis using Miner Lite (Provalis Research) software [35], which is adequate for qualitative analysis. Finally, the codes were presented to peers, refined, and organized as per this report in discussions with other senior researchers from the team.

**Results**

**Usability Questionnaire**

Table 3 shows the questions asked and the number of users who assigned each grade to each criterion. The bot obtained high grades on all evaluation criteria. App usefulness obtained the highest mean (4.57), whereas satisfaction attained the lowest mean (4.38). Figure 3 shows a box plot of the grades assigned by users as per quartile distribution in Table 4, clearly indicating predominance of grades 4 and 5 with few outliers.
Table 3. Grades assigned by users to each question on chatbot usability within a scale of 1 (lowest grade) to 5 (highest grade; n=63).

<table>
<thead>
<tr>
<th>Question</th>
<th>Grade, n</th>
<th>Total, n</th>
<th>Values, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Was this app easy to use?</td>
<td>4 3 1 8 47</td>
<td>63</td>
<td>4.44 (1.16)</td>
</tr>
<tr>
<td>2. Was this app useful to you?</td>
<td>2 0 6 6 47</td>
<td>61</td>
<td>4.57 (0.92)</td>
</tr>
<tr>
<td>3. Was this app satisfactory to use?</td>
<td>4 1 6 6 43</td>
<td>60</td>
<td>4.38 (1.17)</td>
</tr>
<tr>
<td>4. Would you recommend this app to other people?</td>
<td>4 2 2 5 46</td>
<td>59</td>
<td>4.47 (1.16)</td>
</tr>
</tbody>
</table>

Figure 3. Box plot for grades assigned by users on each criterion for assessment.

Table 4. Quartile distribution of grades assigned by users to each question on chatbot usability.

<table>
<thead>
<tr>
<th></th>
<th>Ease of use, n</th>
<th>Usefulness, n</th>
<th>Satisfaction, n</th>
<th>Recommendation, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum value</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Quartile 1 (25%)</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Quartile 2 (50%): median</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Quartile 3 (75%)</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Maximum value</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

Qualitative Assessment

Initially, excerpts were annotated with the following tags: positive feedback (aspects reported as positive by the participants regarding interaction, interface, and content of the chatbot); negative feedback (points considered negative by the participants pertaining to interaction, interface, and content of the chatbot); and neutral (comments that did not qualify as either ostensibly positive or negative). Different themes emerged in each broad category (positive or negative).

As a following step, we analyzed each of the themes and organized them based on whether they were related to the usability of the chatbot or to the health support it offered. Themes associated with the usability of the chatbot pertained to issues related to the system’s interface, the users’ perspective of the effectiveness and efficiency of the proposed functionalities, and users’ perceptions and responses to the use of the system. Health support included all themes that addressed aspects of how the chatbot achieved its goal to offer support regarding COVID-19 screening and education to users.

Our classification allowed us to reveal and point to problems related to different sources in our COVID-19 chatbot—design decisions of the technology itself and how it supports users’ needs for health information in the context of COVID-19 screening and education. Next, we present the results of our analysis and describe each identified theme. We present both the positive and negative aspects that emerged from our analysis. Nonetheless, we examined the negative aspects in more detail, as they point to the aspects that still need to be improved and dealt with in health chatbots.

Positive Feedback

On the basis of our analysis of the positive comments from participants, 6 different themes emerged—3 related to the
usability of the chatbot and the other 3 pertaining to health support.

Regarding the chatbot’s usability (Table 5), participants in general found the interface esthetically pleasing and with good usability (C1). They reported an overall positive experience with the chatbot, mostly because of its ease of use (C2). Finally, some participants showed a reasonable level of understanding about the chatbot’s underlying logic, which is positive in the sense that the interaction improves as the user understands how the technology works (C3).

As for the themes associated with health support (Table 6), participants valued the fact that the screening process was simple and straight to the point, helping users understand the action they should take (C4). Furthermore, they found that there was a broad range of topics in the Q&As, including content related to fake news that had been circulating at the time, and considered the answers concise and easy to understand, a very frequent comment in their interviews (C5). Finally, participants found that the chatbot was useful and valuable, especially considering the circumstances they were living in at the time—it was trustworthy and allowed them to obtain reliable information without the risk of getting infected (C6).

Table 5. Codes, number of occurrences, and examples of positive feedback regarding usability (we translated excerpts into English).

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
<th>Occurrences, n</th>
<th>Examples</th>
</tr>
</thead>
</table>
| C1. Chatbot interface design and functionalities | Comments on chatbot graphic interface design, including font-size and text display on screen, chatbot esthetics, and use of button-limited options | 4 | • “Well, I liked having a menu with numbers and that you can just type a number, because I think it makes it much faster.” [P05]  
• “Spacing in the text display is adequate and the text fits almost all the screen, right? It’s not so small that we have to strain our eyes to read it, and it is also easy to scroll the screen up and down to be able to go back to check something.” [P13] |
| C2. Positive user experience | General positive comments on overall experience of interacting with the chatbot, for example, ease of use | 26 | • “I thought it (the experience) was nice, the bot was very easy to use and you get clear instructions about what you have to do, there is no way you can get lost.” [P12]  
• “It is very good that it is responding very fast. It doesn’t have that, you know, delay to answer.” [P03] |
| C3. Understanding chatbot underlying rationale | Comments on user perception and understanding of rationale behind the chatbot operation | 3 | • “There is an interesting logic underlying the bot. It looks like a flowchart, right? You have a yes or no question, then depending on the answer, you go to another group of questions...” [P01]  
• “...I noticed that depending on the information I submit about my symptoms, it will give me a direction, it will direct me to where to go, isn’t it?” [P15] |
Table 6. Codes, number of occurrences, and examples of positive feedback regarding health support (we translated excerpts into English).

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
<th>Occurrences, n</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>C4.</td>
<td>Patient screening session—process and guidelines</td>
<td>18</td>
<td>• “I thought the bot was very cautious (in its assessment), because I reported having symptoms considered severe, right? So the bot immediately told me to seek help as soon as possible.” [P13]</td>
</tr>
<tr>
<td></td>
<td>Considerations about directions given, color system used in the triage phase, and chatbot guidance during the screening session</td>
<td></td>
<td>• “I thought the guidelines were very clear: options about what you needed to do, if you had any symptoms...like fever...and also questions about you belonging to a risk group, having a risk condition, right? Which could mean a more severe Covid case.” [P01]</td>
</tr>
<tr>
<td>C5.</td>
<td>Question and answer session—range of topics and trustworthiness of information provided</td>
<td>67</td>
<td>• “The content, the number of questions, topics...I was positively impressed. If you had asked me how many general topics I could think of, I wouldn't have thought of 12 at all. You know? That’s what I liked the most, that there’s also a lot of information to be explored.” [P03]</td>
</tr>
<tr>
<td></td>
<td>Number and content of answers considered satisfactory as well as effective in expanding knowledge about disease</td>
<td></td>
<td>• “I thought it [the experience] was very, very positive. I really liked the content ... it is very reliable. These are true guidelines, everything is correct.” [P06]</td>
</tr>
<tr>
<td>C6.</td>
<td>Reported advantages of chatbot use during COVID-19 pandemic</td>
<td>18</td>
<td>• “I would use the bot and recommend it to friends and acquaintances who might want to have some reliable information, because there is plenty of fake news about Covid. And even though it is a robot, you do get a reliable answer; there is no fake information. The bot provides very straightforward answers that point to what should be done.” [P15]</td>
</tr>
<tr>
<td></td>
<td>Motivations and advantages of using a chatbot during the COVID-19 pandemic</td>
<td></td>
<td>• “Here, where I live, this chatbot would be my first option to seek advice, for sure, because I don’t have many options...to get such guidance, to avoid going somewhere where I might get infected, or to speed up my recovery from the disease.” [P08]</td>
</tr>
</tbody>
</table>

Negative Feedback

On the negative side, although we obtained approximately the same number of excerpts as in positive feedback, our analysis led to a larger set of categories, 10 of them related to usability and 5 related to health support.

Regarding usability (Table 7), different types of problems emerged, from technical problems to interaction and interface design problems to problems with the expectation of better communicative capability (which would require artificial intelligence [AI] support). Technical problems were reported by participants who faced difficulties when sharing their location with the chatbot (C7), and sometimes, the app became slow or unresponsive in their mobile phones (C8).

Although participants considered their overall experience with the chatbot to be good, they commented on many issues that could improve the interaction if solved. Regarding the flow of the conversation, some participants had difficulties when trying to go back after typing a wrong option (eg, C9). Some participants complained about the menu being displayed too quickly and hindering their ability to read the chatbot’s (previous) response (C10). Another problem related to conversation flow was observed when participants did not understand how their interaction with the chatbot evolved. For instance, a participant missed the cue indicating that the chatbot was answering and did not wait for it to respond before sending another message (C11).

The lack of graphical interactive resources (eg, clickable menu options) was also an issue for some participants (C12). Another problem we observed in some sessions was participants not knowing how to start the conversation with the chatbot and asking the interviewer for guidance owing to absence of basic initial directions (C13). Some comments about the chatbot language were also pointed out by participants who thought it may not be adequate for users with lower levels of literacy (ie, they may not be able to understand it), which indeed can be an issue in Brazil (C14).
Finally, we also observed some interaction problems related to our chatbot technology limitations. Some participants entered unexpected text inputs into the chatbot that it was not prepared to handle (C15). Similarly, others tried to interact with the chatbot by typing or even speaking in natural language (C16). Both issues could be addressed by applying better support for natural language processing and understanding using AI, which was already commonly found in several conversational systems at the time.

Regarding health support (Table 8), participants commented on some outdated or missing information they noticed in some answers (C17 and C18). It is worth pointing out that interviews took place at the end of 2020 when there was still much to be learned about COVID-19. Furthermore, this was around the time when the vaccine was underway and the chatbot did not have any information about it yet. In some cases, participants reported dismay with the briefness of the clinical evaluation during the screening session (a participant, for instance, expected to handle (C15). Similarly, others tried to interact with the chatbot by typing or even speaking in natural language (C16). Both issues could be addressed by applying better support for natural language processing and understanding using AI, which was already commonly found in several conversational systems at the time.

Table 7. Codes, number of occurrences, and examples of negative feedback regarding usability (see Multimedia Appendix 3 for more examples).

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
<th>Occurrences, n</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>C7</td>
<td>Difficulty in sharing location</td>
<td>Participant unable to share device GPS location when requested by chatbot</td>
<td>10</td>
</tr>
<tr>
<td>C8</td>
<td>Technical problem in mobile app or phone</td>
<td>Chatbot stops responding or gets slow; interaction is interrupted</td>
<td>4</td>
</tr>
<tr>
<td>C9</td>
<td>Need for an option to go back and make a different choice during interaction</td>
<td>Participant needs to start over and repeat entire session, as the chatbot does not have an option for backtracking or choosing a different path during conversation</td>
<td>2</td>
</tr>
<tr>
<td>C10</td>
<td>User choice repeatedly prompted by option menu and at high pace</td>
<td>Participant complaint about being prompted to make a choice in option menu and finding it too fast to be able to read the whole answer provided by the bot</td>
<td>6</td>
</tr>
<tr>
<td>C11</td>
<td>Conversation flow management</td>
<td>Participant does not succeed in keeping the conversation flowing with the chatbot owing to unperceived feedback or lack of it from the chatbot (eg, turn taking management)</td>
<td>4</td>
</tr>
<tr>
<td>C12</td>
<td>Better interface resources</td>
<td>Additional features in chatbot interface to enhance interaction</td>
<td>12</td>
</tr>
<tr>
<td>C13</td>
<td>Insufficient directions on how to interact with the chatbot</td>
<td>Participant requesting directions or help from the interviewer</td>
<td>22</td>
</tr>
<tr>
<td>C14</td>
<td>Chatbot language need to be adapted to meet different user profiles</td>
<td>Language used by the chatbot needs to be adapted to be understood by user with low-literacy level</td>
<td>2</td>
</tr>
<tr>
<td>C15</td>
<td>Chatbot fails to understand unexpected user responses</td>
<td>Chatbot does not successfully process information entered by the participant</td>
<td>11</td>
</tr>
</tbody>
</table>
| C16  | Participants expectations exceed chatbot’s actual communicative ability | Participant tries to interact in a way not supported by the chatbot, for example, by trying to speak to the chatbot by voice | 5 | PO6: So I should type that I would like to know more about pregnancy? [Participant starts typing “I want to know more about pregnancy”]
Interviewer: Actually, each number is a shortcut, you don’t have to type everything. |
a more detailed and thorough evaluation of her symptoms before the chatbot gave her instructions) and the lack of mechanisms to mitigate responses for severe symptoms (C19). Finally, the last themes have to do with the need, mentioned by some participants, for more practical and situated guidance or information both in the screening section (C20) and in the Q&A section (C21).

As is the case with any qualitative analysis, numeric information should be interpreted cautiously and is presented here for the sake of transparency. It should be noted that the number of tagged occurrences of a code is not a general indicator of relevance or importance, because our analysis was not based on frequency or other statistical metrics. Thus, this information is not meaningful to discuss codes’ validity [31] and was included as an index of the overall analysis process, not to indicate any validation of the analysis. Table 9 shows that the number of negative codes is greater than the number of positive codes. This is expected because we analyzed the negative aspects more thoroughly, as stated earlier, leading to individually less frequent and more fine-grained negative codes. In the category level, frequency of codes can be an approximate indicator of the distribution of positive and negative aspects. In Table 9, we can see that 54.8% (136/248) of the excerpts were identified as positive, and the remaining 45.2% (112/248) were identified as negative.

Table 8. Codes, number of occurrences, and examples of positive feedback regarding health support (see Multimedia Appendix 4 for more examples).

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
<th>Occurrences, n</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>C17. Outdated information or answer</td>
<td>Participants noticed some outdated information or questioned whether the information presented in the Q&amp;A session was updated</td>
<td>4</td>
<td>“How often is the FAQ updated? For instance, whether there’s a vaccine or not...Because that ensures credibility, right?...Because sometimes people notice that the information is a little outdated.” [P13]</td>
</tr>
<tr>
<td>C18. Missing information or explanation</td>
<td>Participants suggested a topic in the Q&amp;A session that should be included or further explained</td>
<td>5</td>
<td>“I think there could be some explanation on IgG tests [after reading about IgG tests on one of chatbot’s answers]. Because many people have been talking about it and they don’t know what that is.” [P11]</td>
</tr>
<tr>
<td>C19. Unfulfilled expectations during the screening process</td>
<td>Participants mentioned interesting insights and broken expectations during the screening process</td>
<td>13</td>
<td>“...maybe the person wanted a little bit more information before the chatbot said: ‘Go to the hospital’ [laughs], you know? ‘Go to the hospital because this is a serious symptom.’ Maybe something in the sense of reassuring the person, like...’look, these are symptoms that can be considered...the direction could be modalized, so as not to scare the person.” [P02]</td>
</tr>
<tr>
<td>C20. Need or demand for actionable orientation during the screening session</td>
<td>Participants expected to receive more practical instructions at the end of the screening session</td>
<td>5</td>
<td>“I think there should be something more direct to guide the next step. What am I supposed to do? The bot gave me some explanations about Covid, about my condition, but it didn’t tell me where to go. Given that the person in my scenario is in a risk group, as she’s pregnant, I thought people would need to know about this.” [P06]</td>
</tr>
<tr>
<td>C21. Demand for situation-oriented answers to questions</td>
<td>Participants expected to find answers that could be more directly applied to a particular situation in the question and answer session</td>
<td>7</td>
<td>“If you have traveled, is there a test that allows you to go to your relative’s house without having to worry? Or if you actually have to isolate yourself and wait 3 days to see if you won’t have anything after leaving the airport?” [P09]</td>
</tr>
</tbody>
</table>

Table 9. Summary of code categories.

<table>
<thead>
<tr>
<th>Category</th>
<th>Codes, n</th>
<th>Tagged occurrences (n=248), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Positive</strong></td>
<td>6</td>
<td>136 (54.8)</td>
</tr>
<tr>
<td>User experience</td>
<td>3</td>
<td>33 (13.3)</td>
</tr>
<tr>
<td>Health support</td>
<td>3</td>
<td>103 (41.5)</td>
</tr>
<tr>
<td><strong>Negative</strong></td>
<td>15</td>
<td>112 (45.2)</td>
</tr>
<tr>
<td>User experience</td>
<td>10</td>
<td>78 (31.5)</td>
</tr>
<tr>
<td>Health support</td>
<td>5</td>
<td>34 (13.7)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>21</td>
<td>248 (100)</td>
</tr>
</tbody>
</table>
Discussion

Overall Findings

The WHO guidelines point out that considering the potential impact that interface and interaction issues have on health care services and even on clinical practice [15], it is essential to evaluate user experience in health care systems. Despite the increased use of chatbots in a range of fields, this form of technology has yet to be robustly assessed, and the literature regarding these conversational agents’ formats, focusing on their acceptability, safety, and effectiveness, is still incipient [7]. Moreover, the lack of standardization and paucity of objective measures make it difficult to compare the performance of health chatbots [36].

In this paper, we present users’ evaluations of a chatbot developed specifically for screening cases and supplying information regarding COVID-19. We performed a brief, quantitative assessment with actual chatbot users and an in-depth evaluation with participants through simulated scenarios (volunteers who were asymmetric and engaged in chatbot interactions as guided by the interviewer).

Although our quantitative analysis indicated that overall users were satisfied with the chatbot, our qualitative analysis allowed us to identify participants’ perspectives of positive and negative aspects regarding usability and health support, as described in sections Positive Feedback and Negative Feedback. The positive comments from the qualitative study corroborate the quantitative results we found, as positive comments represented approximately 55% of all comments, and the most frequent codes emphasized an overall positive experience (C1) and the usefulness of the provided health support during the pandemic context (C4-C6). At the same time, the negative comments in the qualitative study are not in conflict with the overall positive experience from the quantitative study. All volunteers from the qualitative study reported having an overall positive experience with our chatbot during the interviews. The negative comments should be interpreted as opportunities for improvement that did not compromise the overall experience. In the subsequent sections, we discuss some of the main issues based on our analysis.

Updated Chatbot Information

The results indicate that the pandemic context created specific circumstances that led participants to assign value to having a chatbot available—fake news dissemination about COVID-19 and the disease’s high transmission rate. This means participants welcomed the possibility of having access to reliable information at a time when plenty of fake news about COVID-19 was circulating in Brazil, presumably connected to political interests and governmental sources as well as misinformation and infodemication from inappropriate scientific papers [37]. Furthermore, knowledge about COVID-19 was rapidly evolving, and the population was seeking sources of trustworthy information. Participants also felt that obtaining directions as to how to proceed in case of symptoms without having to be exposed to chances of getting infected by the virus was a positive factor. On the other hand, because information evolved so quickly, participants noticed that information provided by the chatbot was not fully up to date (eg, about vaccines, which were underway). This was perceived as a negative impact that could undermine the reliability assigned to the chatbot and points to the challenge of the need for constant information updating in conversational agents. This includes deciding which pieces of new information are relevant to be included and how to best translate new scientific evidence for the lay population, a similar challenge faced by decision support systems in general [38]. As previously stated, developing a high-quality COVID-19 chatbot is critical but not enough for widespread adoption. It is fundamental to demonstrate and emphasize that chatbots are able to deliver the same quality service as human agents [39].

Universal Usability

Another aspect that emerged in our analysis, which is very relevant to the Brazilian context and may also be relevant to other resource-limited countries, is the need for universal usability [40], that is, to provide access to technology to all citizens. In the case of our chatbot, issues related to access quality were observed by the interviewer or directly reported by the participants in our evaluation. Some participants used their own cell phones for assessing the chatbot. However, in one case, the participant had technical problems that were not software bugs strictly speaking but seemed to be associated with users’ device limitations related to the operational system and to hardware resources. Although currently there are more smartphones in use in the country than citizens [41], owing to the inequalities in our country, the chatbot may not be universally accessible through all smartphone models in use. Furthermore, one participant specifically raised the issue of the educational level of other Brazilian citizens and pointed out the need to adapt the chatbot’s language to a larger variety of user profiles. These results corroborate not only the need to assess user experience of health care technology in general but also issues brought about by local conditions of technology use in a country or region.

Beside quality of access to technology, literary and accessibility issues are also issues to be tackled in chatbot development. Srivastava [42] reviewed gaps found in using chatbots during COVID-19, and one of them was “inaccessible information,” that is, most of the chatbots created assumed that the users were literate, experienced with digital technology, and did not have any disabilities. These assumptions prevented a considerable part of the society from benefiting from chatbot technology. In the case of our chatbot Ana, our team performed several updates in the chatbot language, aiming at making language more accessible to less-literate users and enhancing user experience.

Expected Communication Abilities

Analyzing the negative aspects of interacting with the chatbot, we noticed that most of them (6 out of 10—C9, C10, C11, C13, C15, and C16) were related to the conversational paradigm adopted in such technologies. Participants reported on interactive breakdowns (ie, problems they had as they interacted with the system) that were generated by many different causes—from not knowing exactly how to interact with it (C13) to expecting too much of the bot’s communicative abilities (C16). Although these challenges are mainly related to chatbots in general [43-47] and not only in the health domain, they emerged as hindering...
users’ interaction with the system and impacted (negatively) their experience with it, which could lead them to not fully embrace or adopt the technology.

Complete Health Care Information
It is important to understand the negative aspects related to the health support offered by the chatbot, as this is the technology’s main goal. Out of the 5 themes describing these negative aspects, 2 of them, as mentioned, were related to the need to keep information updated and complete when knowledge of the disease was continuously evolving during the beginning of the pandemic. A third one was the system not fulfilling users’ expectations regarding a more thorough clinical evaluation or more careful instructions to patients. The decision-making framework of a chatbot is crucial to address this issue. Models based on user-initiated solutions are usually easier to deploy; however, this type of solution may be insufficient in some scenarios and may lead to situations in which a high-risk person or a person with issues regarding specific conditions or contexts would rather seek an in-person assessment in a health care facility, because they did not feel safe or could not follow the recommendations. On the other hand, models based on provider-initiated solutions allow providers to “close the loop” and properly address more specific conditions [11].

Contextual Information and Adaptive Ability
Finally, the last 2 themes (C20 and C21) point to the expectation that some users highlighted of having more practical orientation and situation-oriented information. These kinds of features would demand more sophisticated technologies that might be able to handle contextual information from users to identify contextual needs and adaptively respond to them. To achieve this type of goal, we would need at least a richer data set comprising a reasonable set of different situations and Q&A pairs and more sophisticated technologies able to detect and handle users’ contexts appropriately. Context could be inferred from isolated conversations but would probably be better constructed by technologies that combine external variables (historical data, location, etc) such as search engines and advertisement technologies. We believe that adaptive AI capabilities such as recommender systems can be included in the chatbot to provide more specific instructions that would take into consideration the users’ specific condition (eg, comorbidities) or context (eg, location) in the answers and piece of advice given.

An adaptive approach can also be used as a strategy to address users’ diversity in skills and preferences. We observed conflicts between the participants’ comments and opinions, such as C1 contradicting C9, C10, and C12 and C3 contradicting C13. As mentioned before, most participants in the qualitative study had an overall positive experience with our chatbot, and negative comments should be interpreted as opportunities for improvement and not as a conflict of results. However, the participants were bothered by our chatbot’s problems when interacting with the system in different ways, depending on their profile, background, patience at the time, etc. Fulfilling the goals and needs of a large diversity of users with different profiles, backgrounds, and preferences is also a goal at the core of the universal usability principle [40] and one of the major interaction design challenges. We believe that adaptive chatbots should be investigated as a promising technology to help in this regard.

Directions for Future Research
The construction and deployment of a chatbot for COVID-19 is a dynamic project that demands collaboration among multiple disciplines such as health professionals, linguists, technology designers, and developers [16]. The results of our qualitative analysis and discussions provide directions for multidisciplinary teams to approach projects of prospective bots and are expected to help organize the problem space of regarding interaction decisions and issues to help understand users’ needs and expectations in such endeavors.

Limitations
Although our qualitative evaluation of the chatbot included a small sample of 15 participants, there was a distribution of gender (7 female and 8 male participants) and age, varying from 18 to 62 years. Nonetheless, as the assessment was performed during the pandemic through teleconference, it required participants who had access to computers and good internet bandwidth. Thus, it does not represent the variety of educational or economic groups in Brazil. In the future, our goal is to broaden our evaluation to include other groups of our population who represent potential users of the system. Moreover, we did not investigate the perceptions of physicians, nurses, and caregivers regarding the use of this COVID-19 chatbot, including their benefits, challenges, and risks to patients.

This qualitative study was designed to allow the collection of rich, in-depth data containing participants’ thoughts and insights about their experience of using our chatbot. At that time, using chatbots for health purposes was not common in Brazil, and the interviewer’s clarifications during sessions were given to participants to unblock them from dead ends, thus enabling us to collect more rich and useful data. All such cases were annotated and considered in the analysis.

Our quantitative assessment of our COVID-19 chatbot was evaluated by 10.1% (63/622) of users who chose to participate and expected more users to willingly participate and more data to be collected regarding their attitudes toward the system.

Conclusions
This study evaluated the quality of user experience with a chatbot designed in response to the COVID-19 pandemic by a large telehealth service in Brazil through an analysis of usability with real users and an exploration of strengths and shortcomings of the chatbot, as revealed in reports by participants in simulated scenarios. Our results indicate that overall, users had a positive experience with the chatbot and found the health support relevant. Nonetheless, the qualitative evaluation of the chatbot indicated challenges and directions to be pursued in improving not only our COVID-19 chatbot but also health chatbots in general.
Acknowledgments
The authors thank all the participants of the chatbot assessment as well as chatbot users who took the time to evaluate the chatbot for their contribution to this research. The authors also thank Fernanda Rocha Gonçalves for her contribution to the study design.

This study is supported in part by the Brazilian research agencies Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (grant 88887.516155/2020-00), Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG, grant RED-00081-16), and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)/Instituto de Avaliação de Tecnologias em Saúde (grant 465518/2014-1).

ALPR was supported in part by Brazilian research agencies CNPq (310790/2021-2 and 465518/2014-1) and FAPEMIG (PPM-00428-17 and RED-00081-16). MSM was supported in part by CNPq (grant 310561/2021-3). ASP has a CNPq (313103/2021-6) and a FAPEMIG (APQ-01.461-14) grant; BAC was supported by a Pró-Reitoria de Pesquisa, Universidade Federal de Minas Gerais, Secretaria de Educação Superior, and Ministério da Educação grant (23072.211119/2020-10); TMP received scholarships from Pró-Reitoria de Extensão from the Federal University of Minas Gerais. ALPR and MSM are members of the National Institute of Science and Technology for Health Technology. The sponsors had no role in the design of this study during its execution, analyses, interpretation of the data, or decision to submit the results.

Data Availability
Data are available on reasonable request.

Authors' Contributions
All authors reviewed and edited the manuscript and approved the final version. AB, ASP, BAC, CRAO, ECP, HV, KF, MSM, ROP, and TMP drafted the manuscript. ASP, ALPR, BAC, MSM, ROP, and ZSNR were responsible for the research protocol. ASP, BAC, MSM, and ROP coordinated the study. BAC and ROP were responsible for the application of the questionnaire. BAC, ECP, HV, and KF performed data analysis. CRAO, LBR, MSM, and ZSNR participated in chatbot development and testing. CRAO, LBR, MSM, and ZSNR participated in chatbot implementation.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Scenarios for in-depth chatbot evaluation (asymptomatic users).
[PDF File (Adobe PDF File), 526 KB - humanfactors_v10i1e43135_app1.pdf]

Multimedia Appendix 2
Script for the Interview conducted with asymptomatic (healthy) participants following their interaction with the chatbot.
[PDF File (Adobe PDF File), 702 KB - humanfactors_v10i1e43135_app2.pdf]

Multimedia Appendix 3
Complete version of Table 7.
[PDF File (Adobe PDF File), 576 KB - humanfactors_v10i1e43135_app3.pdf]

Multimedia Appendix 4
Complete version of Table 8.
[PDF File (Adobe PDF File), 574 KB - humanfactors_v10i1e43135_app4.pdf]

References


Abbreviations

AI: artificial intelligence
Q&A: question and answer
RCT: randomized controlled trial
WHO: World Health Organization
©Bruno Azevedo Chagas, Adriana Silvina Pagano, Raquel Oliveira Prates, Elisa Cordeiro Praes, Kícila Ferreguetti, Helena Vaz, Zilma Silveira Nogueira Reis, Leonardo Bonisson Ribeiro, Antonio Luiz Pinho Ribeiro, Thais Marques Pedroso, Aline Beleigoli, Clara Rodrigues Alves Oliveira, Milena Soriano Marcolino. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 03.04.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original work on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Patient Satisfaction With Speech Recognition in the Exam Room: Exploratory Survey

Jeffrey Sippel1†, MD, MPH; Tim Podhajsky1†, MD; Chen-Tan Lin1†, MD
Department of Internal Medicine, University of Colorado School of Medicine, Aurora, CO, United States
†all authors contributed equally

Corresponding Author:
Jeffrey Sippel, MD, MPH
Department of Internal Medicine
University of Colorado School of Medicine
12700 E 19th Avenue
RC2, Room 9C03, Box C272
Aurora, CO, 80045
United States
Phone: 1 303 724 4075
Fax: 1 303 724 6042
Email: jeffrey.sippel@cuanschutz.edu

Abstract

Background: Medical speech recognition technology uses a microphone and computer software to transcribe the spoken word into text and is not typically used in outpatient clinical exam rooms. Patient perceptions regarding speech recognition in the exam room (SRIER) are therefore unknown.

Objective: This study aims to characterize patient perceptions of SRIER by administering a survey to consecutive patients scheduled for acute, chronic, and wellness care in three outpatient clinic sites.

Methods: We used a microphone and medical speech recognition software to complete the “assessment and plan” portion of the after-visit summary in the patient’s presence, immediately printed the after-visit summary, and then administered a 4-question exploratory survey to 65 consecutive patients in internal medicine and pulmonary medicine clinics at an academic medical center and a community family practice clinic in 2021 to characterize patient perceptions of SRIER. All questions were completed by all participants.

Results: When compared to patients’ recollection of usual care (visits with no microphone and an after-visit summary without an “assessment and plan”), 86% (n=56) of respondents agreed or strongly agreed that their provider addressed their concerns better, and 73% (n=48) agreed or strongly agreed that they understood their provider’s advice better. A total of 99% (n=64) of respondents agreed or strongly agreed that a printed after-visit summary including the “assessment and plan” was helpful. By comparing the “agree” and “strongly agree” responses to the neutral responses, we found that patients felt that clinicians using SRIER addressed their concerns better (P<.001), they understood their clinician’s advice better (P<.001), and receiving a paper summary was helpful (P<.001). Patients were likely to recommend a provider using a microphone based on the Net Promoter Score of 58.

Conclusions: This survey suggests patients have a very positive perception of speech recognition use in the exam room.

(JMIR Hum Factors 2023;10:e42739) doi:10.2196/42739

KEYWORDS
speech recognition; exam room; primary care; general practitioner; satisfaction; survey; perception; opinion; speech; voice; eHealth; digital health; health technology; communication technology

Introduction

Health care is increasingly complex due to rising patient severity of illness, electronic health record (EHR) and documentation requirements, and institutional demands to see more patients in a shorter amount of time [1,2]. Simultaneously, clinician burnout is growing due to increases in cognitive workload [1]. The COVID-19 pandemic has further accentuated clinician stress and burnout since 2020 [3]. Taken together, these factors work...
against The Quadruple Aim of health care, which acknowledges the need to improve the work life and professional fulfillment of clinicians [4].

Speech recognition software transcribes the spoken word into text by using a dedicated microphone connected to a computer in conjunction with speech recognition software. Speech recognition has been primarily used to enhance clinician documentation [5]. Although there was dissatisfaction with early versions secondary to time lags and transcription errors, the accuracy and performance of speech recognition have greatly improved [5]. Speech recognition continues to gain popularity in the medical field and has been favorably received by clinicians to improve EHR efficiency [6]. However, speech recognition is not typically used in the clinical exam room in outpatient settings. Patient perceptions regarding the use of speech recognition in the exam room (SRIER) are therefore unknown. If speech recognition were used simultaneously with the EHR in the exam room, a real-time transcription would be available for immediate review. Its use in the exam room to provide a summary of the clinical encounter, such as an “assessment and plan,” may be a patient satisfier by providing a reflective listening opportunity for both patient and clinician, and by documenting the care plan in real time. The transcription can also be printed in the after-visit summary and given to the patient. This report describes our efforts using SRIER and its impact on patient perceptions.

Methods

Study Setting and Participants

We conducted an exploratory survey to examine patient perceptions regarding SRIER. A convenience sample of sequential appointments for acute, chronic, and wellness care was included. We administered surveys to 65 consecutive patients in the fall of 2021. The surveyors included the three authors, and all were attendings who worked in different outpatient clinic settings (internal medicine and pulmonary medicine clinics at an academic medical center and community family practice). Surveys were not administered in inpatient or emergency department settings.

Study Protocol

We used medical speech recognition software and microphone hardware (Dragon Medical One and Dragon Powermic III by Nuance, Burlington, MA) to complete the patient’s “assessment and plan” in their presence in the exam room. This electronically transcribed “assessment and plan” was included in the printed after-visit summary, which was given to all patients immediately upon completion. After the clinic visit, each physician handed a paper survey to the patient. Then the physician left the room. The patient remained in the room to complete the survey, which was collected by a medical assistant. We did not use the Nuance Dragon Ambient eXperience system.

Outcome Measures

The survey was comprised of four questions and comments. We selected three questions that would reflect patient perceptions of SRIER, focusing on the perceived effectiveness of the encounter and the value of a printed after-visit summary including the transcribed “assessment and plan.” We used a 5-point Likert scale for these questions (strongly disagree, disagree, neutral, agree, strongly agree) [7,8]. The last question asked whether the patient would recommend SRIER to others by using the Net Promoter Score (NPS) and asked for any comments [9]. All outpatient exam rooms were equipped with computers and microphone hardware, so there were no operational costs for this survey. This was considered a quality improvement project in the direct care of patients.

Ethical Considerations

The Colorado Investigational Review Board deemed this survey as quality improvement and thus exempt from full review.

Statistical Analysis

One of the authors (JS) collated data from the paper survey results. Descriptive analysis was used to characterize the ordinal variables on the Likert scale questions [7,8]. Standard NPS descriptive analysis was used for the NPS question [9]. To test the null hypothesis (that all participants would be “neutral” to SRIER for the Likert scale questions), we used a 1-sample t test 2-tailed analysis using Analysis ToolPak in Excel (Microsoft Corporation).

Results

All questions were answered by 100% (N=65) of patients (Tables 1-3). Free-text comments were completed by 15% (n=10) of patients. The mean age was 62.8 (SD 12.6) years, and 55% (n=35) were male.

We tested the null hypothesis (that all participants would be “neutral” to SRIER for the Likert scale questions) using a 1-sample t test 2-tailed analysis. By comparing the “agree” and “strongly agree” responses to the neutral responses, we found that patients felt that clinicians using SRIER addressed their concerns better (mean score 4.4 out of 5, SD 0.86; t_{64}=30.17; P<.001), they understood their clinician’s advice better (mean 4.2 out of 5, SD 1.03; t_{64}=9.74; P<.001), and receiving a paper summary that included the transcribed “assessment and plan” was helpful (mean 4.8 out of 5, SD 0.47; t_{64}=12.97; P<.001).
Table 1. Patients’ perceptions of speech recognition in the exam room compared to visits with no microphones.

<table>
<thead>
<tr>
<th></th>
<th>My provider addressed my concerns better (&quot;Agree or Strongly Agree&quot;), n (%)</th>
<th>I understand my provider’s advice better (&quot;Agree or Strongly Agree&quot;), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All sites (N=65)</td>
<td>56 (86)</td>
<td>48 (74)</td>
</tr>
<tr>
<td>Community family practice (n=9)</td>
<td>5 (56)</td>
<td>6 (67)</td>
</tr>
<tr>
<td>Academic internal medicine (n=31)</td>
<td>28 (90)</td>
<td>24 (77)</td>
</tr>
<tr>
<td>Academic pulmonary (n=25)</td>
<td>23 (92)</td>
<td>18 (72)</td>
</tr>
</tbody>
</table>

Table 2. “Did you find it helpful to get a paper printout with what your provider said today?”

<table>
<thead>
<tr>
<th></th>
<th>“Agree or Strongly Agree,” n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All sites (N=65)</td>
<td>64 (98)</td>
</tr>
<tr>
<td>Community family practice (n=9)</td>
<td>9 (100)</td>
</tr>
<tr>
<td>Academic internal medicine (n=31)</td>
<td>31 (100)</td>
</tr>
<tr>
<td>Academic pulmonary (n=25)</td>
<td>24 (96)</td>
</tr>
</tbody>
</table>

Table 3. “How likely are you to recommend a provider using a microphone in the exam room to other patients?”

<table>
<thead>
<tr>
<th></th>
<th>NPSa,b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total (N=65)</td>
<td>58</td>
</tr>
<tr>
<td>Community family practice (n=9)</td>
<td>11</td>
</tr>
<tr>
<td>Academic internal medicine (n=31)</td>
<td>65</td>
</tr>
<tr>
<td>Academic specialty (n=25)</td>
<td>68</td>
</tr>
</tbody>
</table>

aNPS: Net Promoter Score.
bNPS ranges from –100 to 100.

Discussion

Principal Findings

Patients prefer speech recognition by their physician in the exam room compared to their recollection of usual care. Usual care does not include the use of a microphone with speech recognition and does not include a live narrative summary of the visit or routine printing of the after-visit summary. The process of listening to the physician verbally summarize their visit, then receiving a copy of this in their printed after-visit summary was rated positively, leading to feeling heard and understanding medical advice better. This exploratory survey supports speech recognition use in the exam room and suggests that SRIER can enhance physician-patient communication. Representative free-text comments from patients included the following:

Super helpful with the recap and microphone, helps me to ask any question in case I forget something. The microphone allowed me to hear and read a second summary of my issue and treatment.

We observed that the community family practice clinic, although still receiving positive scores for the questions “provider addressed my concerns better” and “I understand my provider’s advice better,” did have scores lower than the academic internal medicine and pulmonary medicine clinics. This may be due to the small sample size. The community family practice clinic also had shorter appointment times than the other clinics, which may have contributed to the difference (20 minutes vs 30 minutes for established patients, and 40 minutes vs 60 minutes for new patients).

Comparison With Other Work

This is the first report of speech recognition use in the exam room that we are aware of. The physician’s workflow included verbally summarizing the patient’s concerns and stating the “assessment and plan” in real time in the patient’s presence. The physician’s statements were captured as part of the EHR by speech recognition and printed in the after-visit summary. This workflow allows the provider to attend to the patient rather than typing notes into the computer, which improves patient comprehension and physician understanding. Reflective listening is a technique where the clinician repeats some of the patient’s words to indicate understanding. The use of speech recognition allows such word repetition to be documented in the note and serve a similar purpose. Verbally stating the “assessment and plan” allows patients to ask clarifying questions and correct misunderstandings. Since this portion of the documentation is completed in real time, it may reduce “pajama time” and the risk of burnout from after-hours work [6]. The real-time workflow also improves documentation accuracy by not relying on memory recall to complete notes hours or days later and reduces duplicate documentation. The printed summary reduces the risk of patients forgetting unwritten advice, which can be as high as 40%, and allows them to share advice with others [10]. These findings are relevant to acute, chronic, and wellness
visits, and were not assessed in either the inpatient or emergency department settings.

Limitations
There are several limitations to this exploratory survey. The project was conducted by three physicians at one health care organization with a limited set of patients, so the findings are not intended to be generalizable. The survey was conducted in outpatient clinic exam rooms, not in the inpatient or emergency department settings. Survey questions were not validated. Physician satisfaction was not assessed due to the small number of participants. There was not a control group for this survey, since the questions asked the patients to reflect and compare the current visit using speech recognition to their prior clinic visit experiences without speech recognition, essentially providing their own control. The accuracy of patient recall for satisfaction with prior visits was not validated.

Future investigation should expand on both patient and clinician experience with speech recognition in exam rooms. We are aware that developing technologies may capture full conversations between clinician and patient to auto-generate progress notes. It is clear that patients may be ready for such automation tools based on our initial findings.

Conclusion
Patients have a very positive perception of speech recognition when used in the exam room. Periodic assessments such as this will be helpful to understand patient perceptions more fully as the use of technology by clinicians continues to change and expand. As speech recognition technology improves, similar surveys of patients and clinicians can guide the optimal use of such tools to improve communication, improve care, and reduce documentation burden.

Conflicts of Interest
None declared.

References

Abbreviations
- EHR: electronic health record
- NPS: Net Promoter Score
- SRIER: speech recognition in the exam room
Detecting Medication-Taking Gestures Using Machine Learning and Accelerometer Data Collected via Smartwatch Technology: Instrument Validation Study

Chrisogonas Odero Odhiambo¹*, BSc, MS, PhD; Lukacs Ablonczy²*, BSc; Pamela J Wright³*, CEN, RN, MS, MEd, PhD; Cynthia F Corbett³*, RN, PhD; Sydney Reichardt²*; Homayoun Valafar¹*, BSc, MS, PhD

¹Department of Computer Science and Engineering, University of South Carolina, Columbia, SC, United States
²Honors College, University of South Carolina, Columbia, SC, United States
³Advancing Chronic Care Outcomes through Research and Innovation Center, College of Nursing, University of South Carolina, Columbia, SC, United States
*all authors contributed equally

Corresponding Author:
Homayoun Valafar, BSc, MS, PhD
Department of Computer Science and Engineering
University of South Carolina
315 Main Street
Columbia, SC, 29208
United States
Phone: 1 8037774046
Email: homayoun@cec.sc.edu

Abstract

Background: Medication adherence is a global public health challenge, as only approximately 50% of people adhere to their medication regimens. Medication reminders have shown promising results in terms of promoting medication adherence. However, practical mechanisms to determine whether a medication has been taken or not, once people are reminded, remain elusive. Emerging smartwatch technology may more objectively, unobtrusively, and automatically detect medication taking than currently available methods.

Objective: This study aimed to examine the feasibility of detecting natural medication-taking gestures using smartwatches.

Methods: A convenience sample (N=28) was recruited using the snowball sampling method. During data collection, each participant recorded at least 5 protocol-guided (scripted) medication-taking events and at least 10 natural instances of medication-taking events per day for 5 days. Using a smartwatch, the accelerometer data were recorded for each session at a sampling rate of 25 Hz. The raw recordings were scrutinized by a team member to validate the accuracy of the self-reports. The validated data were used to train an artificial neural network (ANN) to detect a medication-taking event. The training and testing data included previously recorded accelerometer data from smoking, eating, and jogging activities in addition to the medication-taking data recorded in this study. The accuracy of the model to identify medication taking was evaluated by comparing the ANN’s output with the actual output.

Results: Most (n=20, 71%) of the 28 study participants were college students and aged 20 to 56 years. Most individuals were Asian (n=12, 43%) or White (n=12, 43%), single (n=24, 86%), and right-hand dominant (n=23, 82%). In total, 2800 medication-taking gestures (n=1400, 50% natural plus n=1400, 50% scripted gestures) were used to train the network. During the testing session, 560 natural medication-taking events that were not previously presented to the ANN were used to assess the network. The accuracy, precision, and recall were calculated to confirm the performance of the network. The trained ANN exhibited an average true-positive and true-negative performance of 96.5% and 94.5%, respectively. The network exhibited <5% error in the incorrect classification of medication-taking gestures.

Conclusions: Smartwatch technology may provide an accurate, nonintrusive means of monitoring complex human behaviors such as natural medication-taking gestures. Future research is warranted to evaluate the efficacy of using modern sensing devices and machine learning algorithms to monitor medication-taking behavior and improve medication adherence.

(JMIR Hum Factors 2023;10:e42714) doi:10.2196/42714
KEYWORDS
machine learning; neural networks; automated pattern recognition; medication adherence; ecological momentary assessment; digital signal processing; digital biomarkers

Introduction

Background
Over 3 decades of international research has indicated that complete models of human health comprise complex interactions of biological, behavioral, and environmental factors. While substantial technological advances exist in the study of the biological and environmental bases of diseases, there have been relatively minor advances in technologies for characterizing human behaviors that influence health. Technological devices have pervaded and revolutionized much of our social and private lives, yet their implementation and use in health care remain sparse. In particular, the innovative use of existing, widely used, and commercially available technologies to influence health-promoting behaviors has been underexplored. Adapting smart technologies, such as phones and watches, has the potential to initiate more effective health-promoting interventions for behaviors such as weight loss, physical activity, and medication adherence. Adapting these devices to promote healthier behavior requires solving the crucial problem of characterizing and monitoring human behavior in a way that will be useful, unobtrusive, and personally relevant. Once resolved, the subsequent steps in developing optimal and personalized interventions can be explored.

Better understanding of behavioral activities such as eating, smoking, sleeping, exercising, and medication taking can have a substantial impact on population and individual health, with the potential to significantly reduce overall health care costs worldwide. In this study, we focused on the global public health challenge of medication adherence, as only approximately 50% of people adhere to their medication regimens [1]. Medication adherence, defined as taking medicines according to decisions agreed upon between prescribing health care professionals and patients [2,3], is a complex human behavior critical for the management of chronic health conditions. Studies have identified forgetfulness as the main reason for nonadherence to many long-term medicines [4]. To address forgetfulness, findings from a meta-analysis of medication adherence interventions among adults demonstrated that linking medication taking with existing daily routines and using behavioral strategies (eg, prompts to take medication) are the most effective approaches to promote adherence [4]. Smartphone apps and other technology-based reminders have also shown promising results in promoting medication adherence [5-8]. However, practical mechanisms to determine whether a medication has been taken or not, once people are reminded, remain elusive.

Different methods, both direct and indirect, exist to measure whether a medication has been taken. However, none are considered a gold standard. Direct measurements, such as clinical biomarker specimens or metabolites from pharmaceutical metabolism and direct observations of medication taking, can be expensive and impractical, especially in large population settings [9]. Indirect methods, such as pill counts, electronic monitoring, and self-reporting, offer simpler alternatives but, at best, approximate adherence through proxy data that can be initially overestimated with even less reliability over time [10]. An ideal method to measure medication adherence should be accurate, affordable, and practical (ie, easy to implement).

Recent advances in sensor technology and artificial intelligence (AI) present an innovative opportunity to measure medication adherence objectively, unobtrusively, and conveniently. Wearable devices such as smartwatches may offer the platform to observe medication adherence [11-13]. From a modest 37 million units in 2016, smartwatch shipments worldwide are projected to increase by 253 million units by 2025 [14]. Smartwatches are likely to become as pervasive as cell phones, as their prices continue to decline and they become more advanced with additional sensors and mobile health applications.

Anticipating this trend, our team investigated smartwatch use not only for medication reminders but also as a strategy for monitoring medication adherence. In this report, we present an artificial neural network (ANN) approach [15] that can detect the complex behavior of medication taking, called the natural medication-taking event (nMTE), with as high as 95% accuracy using sensor data available from common smartwatches.

Previous Work
The use of sensors to automatically detect human activities was pioneered by the work of the Neural Network house and was reported in the late 1990s [16-19]. Several studies have illustrated how smartwatches can be used to monitor and detect human motions of interest, such as smoking. [20-23] or falls among older adults [24-27]. Independent reports [13,28-32] have also confirmed the usability of smartwatches to detect other human motion behaviors, such as eating, physical activity, and foot movement. In the last decade, inspired by the introduction of smart wearable devices, human activity recognition has expanded to include activities such as cigarette smoking [20,22], falls [33-35], and sleep [36,37]. Sleep activity has been studied further using sensor data obtained from electroencephalograms and electromyogram devices to develop neural network models [38]. To detect medication taking, numerous approaches and technologies have been introduced, including experimental devices worn on wrists [15,39-41], sensors worn around the neck to detect swallowing [42-44], and vision modules embedded in smart environments such as Microsoft’s EasyLiving project [45,46]. The EasyLiving project showcased the early investigations into context-aware computing using an array of video-capture devices instead of more traditional physical sensors. By using several vision modules in each room, the system could identify motions, people, gestures, and the surrounding environment. The project also focused on the geometric relationships between people, places, and things to build context and form interaction information that would associate objects with their likely use, which could later be used in a more intelligent system for behavior prediction.

https://humanfactors.jmir.org/2023/1/e42714
Although vision-based medication adherence monitoring is a viable human activity recognition method, users’ privacy concerns and the identifiable nature of the data were strong deterrents to the adoption of this method. In contrast, sensor-based smartwatches provide a scalable and practical platform for convenient, unobtrusive, and secure study of human behavior in natural settings (e.g., people’s homes). Our previous work highlighted the potential of smartwatches to monitor medication-taking events (MTEs) under protocol-guided (scripted) conditions (scripted MTEs [sMTEs]), where all participants followed the same method of taking their medication (e.g., use of the right hand to perform most activities) [15]; however, an nMTE may significantly depart from the scripted method. For example, a person may prefer to take the pill with their right hand while drinking water with their left hand. To establish a more practical application of this technology, we explored the feasibility of detecting unscripted events (nMTEs), which extends our previous study [15]. The nMTE may be more generalizable, and thus, it is a more powerful approach to accurately monitor medication taking and measure medication adherence. The nMTE model has the potential to be an effective intervention tool that can increase adherence, reduce accidental instances of overmedication, and be used for medication adherence monitoring by support persons or health professionals.

**Objective**

The purpose of our study was to test the capabilities of our detection model; we used sensor data from MTEs (sMTEs and nMTEs), other similar activities (e.g., eating and smoking), and a dissimilar activity of jogging.

**Methods**

**Participant Recruitment and Data Collection Process**

**Overview**

The study was conducted during the height of the COVID-19 pandemic and required substantial departure from the traditional means of engaging human participants in sensor recognition studies, which have primarily occurred in laboratory settings. The participants (N=28) were recruited using snowball sampling. The inclusion criteria were adults willing and able to complete the study protocol following training. The exclusion criteria were any type of movement disorder (e.g., Parkinson disease) or paresis (e.g., muscular weakness owing to conditions such as stroke). An appointment was made with each potential participant to explain the purpose, benefits, and risks of the study and to address any questions or concerns. After obtaining informed consent, the participants completed a demographic questionnaire and then received a packet with a smartwatch and phone, 2 charger cables, a user manual, a pill bottle, and placebo medication. The user manual was presented and discussed in detail to the participants. Before data collection, the research team members had an internet-based meeting with each participant to train them to collect and transfer the data, which culminated in participants properly demonstrating the activities.

**Collected Data**

Figure 1 depicts the smartphone and smartwatch with triaxial sketches. The smartwatch was used to collect data, and the phone was used to upload the data to cloud storage. The participants wore the watch on their right wrist for sMTEs and on their wrist of preference for nMTEs. The collected data contained hand-motion accelerometer sensor logs of the triaxial values recorded by the watch at a sampling rate of 25 Hz. The data included the time stamp, orientation, and acceleration of the hand during medication-taking activities. The xyz-sensor values were logged into a CSV file by the medication-taking app on the watch. The file was periodically and asynchronously moved to the phone via Bluetooth.

**Data Collection Protocol**

When the participants received their study supply packet, we collected demographic data and scheduled a protocol training session. Following the medication-taking training sessions, the participants independently completed the data collection in their homes. The exercise comprised the first week (i.e., 5 days) of performing medication-taking behaviors using the participant’s natural way of taking medications (nMTEs) and the second week of performing medication-taking behaviors according to a scripted protocol (sMTEs) [15]. Each participant was directed to record 10 nMTE gestures per day for the first 5 days and then 10 sMTE gestures for the next 5 days. In total, 1400 nMTEs and 1400 sMTEs were collected, tallying 2800 gestures.
To enable seamless transfer of data from the watches to cloud storage, each watch was paired with a smartphone. A custom Android application called MedSensor, a software developed by the research team, was installed on both the watch and phone. At the participants' convenience, the watch data were transferred to the phone via Bluetooth connectivity.

**Closure**

After collecting and transferring 10 days of collected data, the participants returned the smartwatch and phone to the study project coordinator and received a US $25 gift card as an incentive. The smart devices were sanitized according to the Centers for Disease Control and Prevention (CDC) guidelines before use by the other participants.

**Ethical Considerations**

The study was reviewed by the University of South Carolina Institutional Review Board and received exemption from Human Research Subject Regulations (Pro00101203). Potential participants were informed of the study purposes, potential risks and benefits, and their rights as research participants, including the voluntary nature of participation. All participants provided verbal consent before study participation.

**Data Preparation and Annotation**

**Overview**

The proper use of data in supervised machine learning (ML) approaches requires reliable annotation of the data. The process required a supervisor (an expert) to identify and define the gestures of interest to be used during the training of ML models. To develop an understanding of which signal constitutes an MTE, in our previous work, we used the sMTE data collected to understand the individual components of the medication-taking gesture. The exact details of the scripted gesture can be found in our previous report [15]. Using this information, the team members identified and annotated the individual gestures of the nMTEs. The process of gesture identification and annotation was accelerated by having participants self-report MTEs on their smartwatches that included time stamps indicating the beginning and ending of each MTE.

The raw data files logged at the cloud repository consisted of a time stamp that included hours, minutes, seconds, and milliseconds; a date that included day, month, and year; and the x, y, and z components of the accelerometer data. The second file contained the time stamps corresponding to the start and end of each MTE reported by the participant. In theory, the self-reported MTE should be sufficient to identify the gesture of interest (ie, medication taking). However, in practice, participants may erroneously report the activity, or the time stamps may only approximate the correct start and end times of the activity. Therefore, each gesture is visually confirmed to ensure high-quality data. A separate utility program was developed to facilitate this process and to create the final usable data [47]. After this final step, the data files were presented in a usable format for training and testing the AI model.

**Secondary Data Acquisition and Preparation**

This study integrated accelerometer data from 4 different human activities (Table 1), with the primary focus on recognizing nMTEs as recorded by smartwatches (Polar M600, Asus Zenwatch, Motorola, and TicWatch), as described in the previous section. The sMTE data were recorded by each participant wearing the smartwatch on the right wrist, then sequentially performing the mini-activities of medication taking (opening the bottle, dispensing pills to the right palm, placing pills in the mouth with the right hand, drinking water with the left hand, and closing the bottle). For the nMTE, the participant performed the same mini-activities in any sequence that is natural to them (ie, how they would typically take medicine). Smartwatch data for other behaviors (ie, smoking and eating) consisted of data reported in previous studies [48,49], whereas the MTE data were collected in this study using the protocol described in the previous section. The jogging data set, by contrast, is open public data from the Wireless Sensor Data Mining Laboratory [50] that were recorded using a smartphone strapped to the waist of the participant.

**Table 1. Summary of all the data sets used in the study.**

<table>
<thead>
<tr>
<th>Activity</th>
<th>Data points, n</th>
<th>Gestures, n</th>
<th>Participants, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medication</td>
<td>824,000</td>
<td>2800</td>
<td>28</td>
</tr>
<tr>
<td>Eating</td>
<td>272,822</td>
<td>5434</td>
<td>6</td>
</tr>
<tr>
<td>Smoking</td>
<td>62,823</td>
<td>1279</td>
<td>12</td>
</tr>
<tr>
<td>Jogging</td>
<td>287,461</td>
<td>5883</td>
<td>27</td>
</tr>
</tbody>
</table>

**Data Preprocessing and Standardization**

Before integrating data from multiple studies, several normalization and standardization steps were performed. Specifically, attention was paid to the consistent standardization of the accelerometer data and sampling rate of the data. Because the devices used for data collection across all studies were Android devices (vs Apple devices), the sensor data followed a common frame of the x-, y-, and z-axes. As the next step, all data sets were processed to adhere to a sampling rate of 25 Hz by excluding data points (in the case of oversampling) or resampling based on the interpolation of the data (in the case of undersampling). To normalize for the different numbers of gestures per activity, the individual gestures were represented multiple times in our data set to provide a balanced representation of activities.
Development of the ANN

Neural Network Platform and Architecture

The human activities of interest in this study—medication taking, eating (pizza), smoking, and jogging—each included a sequence of mini-activities that have temporal properties. The temporal property is a crucial gesture sequence component in the complex activity recognition. For example, the MTE comprised a series of mini-activities, namely (1) open-bottle and dispense-medicine; (2) hand-to-mouth, pill-into-mouth, and hand-off-mouth; and (3) pick-up-water, drink-water, lower-cup-to-table, and close-bottle. The temporal property (time stamp in this case) determines the sequence of the mini-activities, thereby determining the uniqueness of each complex activity. Considered through the lenses of a linguist, where the mini-activities form the words and the time stamp is the order of activities, the meaning or semantics of the full activity was determined by the syntax of this activity sentence [51]. On the basis of the architecture that incorporates long-term memory, the long short-term memory (LSTM) neural network is the best candidate for the implementation of human activity recognition systems. An LSTM neural network is an artificial recurrent neural network (RNN) architecture with feedback connections that facilitate awareness of past activities at the present time of the activity [52,53]. Figure 2 illustrates a typical LSTM cell where \( x_t \) is the input vector to the LSTM unit; \( h_t \) is the hidden state vector (or LSTM unit output vector); \( c_t \) is the cell state vector; and \( c_{t-1} \) is the cell input activation vector. In this study, our model contained 2 fully connected and 2 LSTM layers (stacked on each other), with 64 units each. The learning rate was set at 0.0025.

![Figure 2.](https://humanfactors.jmir.org/2023/1/e42714) The long short-term memory cell can process data sequentially and keep its hidden state through time (reproduced from Chevalier [54], which is published under Creative Commons Attribution 4.0 International License [55]). c: memory cell; h: hidden state or output vector of the LSTM unit; x: input vector to the LSTM unit; tanh: activation function; subscript t indexes the time step.

Training and Testing Procedure

The LSTM network was trained for 150 epochs using the annotated data while keeping track of accuracy and error. Network training is a process in which a ML algorithm is fed sufficient training data to learn from. The training requires multiple passes on the training data. Epoch refers to the total number of iterations of all the training data in one cycle for training the ML model. The batch size was maintained at 1024; batch size refers to the number of input samples that are passed to the network at once. The train and test data sets were partitioned in an 80:20 ratio, respectively, after the balancing procedure. We applied L2 regularization (Ridge Regression) to the model. The L2 penalty/force removes a small percentage of weights at each iteration, ensuring that weights never become zero. Consequently, the penalty reduces the chance of model overfitting.

Each recording data session may contain hours or days of sensor data. The immediate question to answer is what portion of this recording should be presented to the neural network to identify an activity. The LSTM network expects the training data input of fixed length, also referred to as the “window size.” In this study, a window size of 150 points (ie, 150 rows of sensor data logs) was empirically determined to provide an acceptable performance. At a sampling frequency of 25 Hz, this translates to approximately 1.5 seconds of contiguous recorded data. While the window size represents the portion of the raw data that should be in direct view of the ANN at the time of classification, any relevant past contextual information is saved in the internal cell of the LSTM architecture. The temporal exposure of the LSTM-RNN can be accomplished in various ways, including a sliding window of appropriate size. In this study, the sliding window size of 150 points was selected as the optimal compromise between the performance, simplicity, and responsiveness. The sliding window with overlap significantly transforms and reduces the training data set. Furthermore, the transformation assigns the most common activity (ie, mode) in the exposed window of 150 points as a label for the sequence. This is necessary because some windows may contain ≥2 activities, but the mode is considered to be the dominant or overriding activity. Consequential to the input definition, the data were reshaped into sequences of 150 rows, each containing \( x, y, \) and \( z \) values with 10 points of overlap between 2 consecutive windows. The desired output of the system was based on one-hot encoding of the labels to transform them into numeric values that can be processed by the model [56,57].
**Evaluation of the Trained Network**

During the training phase of an ANN, a single metric of performance needs to be defined to assess the network performance. The network performance metric is used by the operator to direct the network and improve the overall performance. In this study, we evaluated the performance of the classifiers using the accuracy metric defined in the equation 1. In equation 1, true-positive (TP) results represent the correctly classified positive examples; true-negative (TN) results represent the correctly classified negative examples; false-positive (FP) results represent negatives misclassified as positives; and false-negative (FN) results represent positives misclassified as false.

\[
\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (1)
\]

As a network performance measure, the accuracy does not account for the bias arising from unbalanced data sets. To remove the effect of unbalanced data (unequal representation of different activities), data within each activity were repeated to arrive at an approximately equal number of representations for medication taking, eating, smoking, and jogging. Despite the adjustments to enforce data set balance, some data sets remained larger than others, potentially translating into a biased favor for the majority classes. Therefore, the study team considered the following additional evaluation criteria: precision, recall, F-measure, and specificity. Precision indicates the fraction of positive predictions that are truly positive. Recall (positive) indicates the fraction of all positive samples that are correctly predicted as positive by the classifier (TP rate). Recall (negative) indicates the fraction of all negative samples that are correctly predicted as negative by the classifier (TN rate).

\[
\text{Precision} = \frac{TP}{(TP + FP)} \quad (2)
\]
\[
\text{Recall} = TP / (TP + FN) \quad (3)
\]

F-measure is the combination of precision and recall. This is calculated as follows:

\[
F\text{-measure} = \frac{([1+ \beta^2] \times \text{recall} \times \text{precision})}{(\beta^2 \times \text{recall} + \text{precision})} \quad (4)
\]

where \(\beta\) is a weighting factor and a positive real number. It is used to control the importance of recall and precision.

\[
\text{Specificity} = \frac{TN}{(TN + FP)} \quad (5)
\]

**Results**

**Sample Demographics**

Recruited participants (N=28) had a mean age of 27.25 (range 20-56) years, and 57% (16/28) were male. The majority were college students (21/28, 71.4%), single (24/28, 86%), and working at least part-time (17/28, 61%). The sample represented racial diversity with 43% (12/28) Asian and 43% (12/28) White individuals, 10% (3/28) belonging to ≥2 races, and 4% (1/28) African American individuals [58]. Most participants (23/28, 82%) were right-hand dominant, whereas only 1 (4%) participant was ambidextrous.

**Visualization of Medication-Taking Protocol Gesture**

As the first step in performing activity recognition with wearable devices, a more detailed understanding of the gesture of interest needed to be developed. Figure 3 represents an example of an entire sMTEs recorded from a right-hand dominant participant. After careful and repeated examination of the gesture, sequential segments of the gesture were identified (Figure 3). When considering the portion of the gesture corresponding to water drinking (phase C), the gradual increase in the accelerometer’s y-axis depicts the beginning of the drinking phase. It can be used both as a hallmark of an MTE and for quantifying drinking duration.

Our medication-taking study consisted of sMTEs and nMTEs. We based the scripted protocol on the natural behavior observed in most of our preliminary studies. Nevertheless, people’s nMTEs varied from the sMTEs as illustrated in Figures 4A, 4B, and 4C.

The visualizations in Figures 5-8 represent the unique signatures of the following activities: pizza eating, medication taking, smoking, and jogging activities.

It is important to highlight the challenging task of identifying nMTEs, given the gesture diversity in mini-activities among different participants. For instance, the simple method of drinking water between 2 participants can vary significantly as illustrated in Figures 4B and 4C. The participant in the Figure 4B performed the task of drinking water with a sudden removal of glass from the mouth, and the participant in Figure 4C removed the glass from their mouth more gradually. These differences in the individual mini-activities are the root of the challenges associated with smartwatch gesture detection. Nevertheless, the combined activities of medication taking are distinctive from the other activities as illustrated in Figures 5-8.
Figure 3. Atomic segmentations of the complex activity, the medication-taking event, corresponding to (A) open-bottle and dispense-medicine, (B) hand-to-mouth, pill-into-mouth and hand-off-mouth, (C) pick-up-water, drink-water, lower-cup-to-table and close-bottle, and (D) after the medication.

Figure 4. Illustration of medicine-taking protocol differences between (A) scripted gesture from user1, (B) natural gesture from user1, and (C) natural gesture from user2. The 3 recorded sessions illustrate the diversity in natural gestures and potential departure from the scripted gestures.

Figure 5. Three bites of pizza-eating activity.

Figure 6. Single medication-taking event.
Validation and Annotation of MTEs

The study of human behavior using wearable devices has several advantages over the traditional self-report methods. Specific to medication taking, self-reported adherence is known to be overestimated [59]. In comparison to self-reports, wearable devices can provide additional useful information such as the time of the day the medication was administered and the number of times the medication was taken in a day without incurring additional time, effort, or cost to the user. The natural gestures duration ranged between 5 and 331 seconds, with a mean of 18.47 (SD 14.34; median 17) seconds per gesture. The scripted gestures duration ranged between 4 and 686 seconds, with a mean of 20.11 (SD 20.65; median 18) seconds per gesture. Considering the average time needed to complete an MTE, outliers can be examined. We considered gestures longer than 100 seconds or shorter than 8 seconds as outliers. There were 6 scripted gestures and 2 natural gestures of a duration of ≥100 seconds. There were also 63 scripted gestures and 40 natural gestures of a duration of ≤8 seconds. In this study, we considered outliers (both natural and scripted) as gestures with a duration of ≤8 seconds for the lower category or ≥100 seconds for the upper category. To determine the outliers, we considered the mean and SD for natural gestures at 18.47 (SD 14.34) seconds and scripted gestures at 20.11 (SD 20.65) seconds. Fewer outliers were observed in the sMTEs than in nMTEs. For the lower category, a random sample of 20 out of 103 gestures was examined, and all were invalid gestures, indicating that the users probably annotated the start/stop gestures in quick succession. By contrast, for outliers of ≥100 seconds in duration, most contained ≥1 MTE gestures in most cases. To better understand the cause of these temporal discrepancies and therefore, validate or invalidate the reported gestures, each recording session was examined by a team member for validity. The results are presented in Table 2.
Table 2. Analysis of upper-category outliers in seconds.

<table>
<thead>
<tr>
<th>Participants</th>
<th>Duration (seconds)</th>
<th>Observations</th>
<th>Correction</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>371</td>
<td>The participant reported 7 consecutively taken medications as 1 medication event.</td>
<td>Individual MTP(^a) events were separated by an ML(^b) supervisor.</td>
</tr>
<tr>
<td>1, 2, 3, 4, 5, and 6</td>
<td>162</td>
<td>One MTP event was observed with some unrelated activities at the beginning or end of the recording.</td>
<td>The unrelated portions of the recording were trimmed.</td>
</tr>
<tr>
<td>3</td>
<td>279</td>
<td>Comprises random activities that do not match medication-taking gesture pattern</td>
<td>__(^c)</td>
</tr>
</tbody>
</table>

\(^a\)MTP: medication-taking protocol.
\(^b\)ML: machine learning.
\(^c\)Significant aspect of human gestures (that in simulation of human gestures, participants are still bound to perform other random activities, probably out of interruption or disruption). This emphasizes the natural medication exercises where every activity happens in the context of other activities.

ANN Training and Testing

As an initial step in the training of a neural network, examining the learning curve can be instrumental. Figure 9 illustrates the learning curve of the designed LSTM-RNN after the proper treatment of the outlier data. The figure shows the training accuracy (depicted in green) for the training and testing sets (dashed and solid lines, respectively). Here, the consistently increasing values of accuracy is an indication that the network is successfully learning the classification task. The agreement of accuracies reported during the training and testing data sets indicates that the network is successful in generalizing the problem and not performing memorization of the training patterns (avoiding overfitting). The patterns shown in red describe the error functions for the training and testing data sets (dashed and solid lines, respectively). A decreasing value of the error function is a further indication of successful learning, with a gradually plateauing pattern that indicates a saturated training session. Table 3 summarizes the performance metrics for the final trained neural network that used a fixed window size of 150 units. The accuracy, precision, recall, \(F\)-measure, and specificity, as described by the equations 1, 2, 3, 4, and 5, respectively, are presented in Table 3. The test accuracies for eating, jogging, medication, and smoking were 94.3%, 100%, 93.6%, and 98.6%, respectively. The average performance attained was 96.6%. To explore the full nature of misclassifications, the confusion table was examined (Table 4).

Figure 9. Training plot for the window size of 150 units. The configuration resulted in the best performance among the different models.
Discussion

Principal Findings

This study aimed to examine the feasibility of detecting nMTEs using smartwatches. Studying human activities using smart and wearable devices has numerous advantages over the traditional approaches. Wearable devices provide the advantage of unobtrusively, continuously observing human behavior in their natural settings, with little burden on the user. The collected sensor data from these devices can be used to validate the data reported by the user, thereby improving the accuracy and completeness of the self-reports. In this study, participants used the smartwatch to report the beginning and end of their MTEs. Errors in the participants’ self-reported MTEs were identified. Some self-reported MTEs had implausibly short or long durations. By validating the digitally recorded temporal gestures, we demonstrated the ability to correct erroneous self-reports, thereby improving the quality of the reports. Furthermore, the temporal signature that has been reported by the array of sensors available on wearable devices can provide a plethora of additional information such as the temporal variation of an activity for a given user or across a population of users. For example, in our study, we demonstrated that nMTEs were completed in an average of 18 seconds, but there were distinct differences across participants. Such a comparison of behaviors provides several dimensions along which the study of human behavior can be expanded.

While the ANN’s overall best performance of 96.6% accuracy in identifying MTEs from other activities was good, there are other nuanced configurations that can be explored. In addition to the sliding window size, it is possible to manipulate the hyperparameters, such as the learning rate, number of LSTM units, window step sizes, and batch sizes, to achieve an even better performance. We will consider these in future experiments. In addition to the expanded information that can be obtained from these devices, the ability to automatically detect and identify an nMTE with high accuracy will be beneficial. The automatic detection of an MTE event can be used as the foundation for both monitoring MTEs and improving medication adherence. In the latter case, the nondetection of an nMTE offers the opportunity to alert patients or support persons who missed medications to improve adherence and the health outcomes associated with improved medication adherence. Improved medication adherence has the potential to significantly reduce morbidity [60-62], mortality [3,62], and health care costs [60,63-65]. Hence, detecting nMTEs using smartwatches has exponential utility.

This study addressed several gaps and limitations of other studies. Fozoonmayeh et al [66] used Android LG Watch Sport smartwatches with cellular capability, eliminating the need for a smartphone to transfer data to the cloud repository. The study also relied on both the accelerometer and gyroscope sensor data to detect motions. Data collection and subsequent transmission to the cloud storage were completed in real time, requiring the watch to have constant data connectivity. Constant connectivity uses large amounts of energy; therefore, battery life may be problematic and data integrity can be compromised by poor or noisy connectivity. Our study simplified data collection by adopting an offline approach. However, it relies on a paired smartphone to relay the data to cloud storage. Importantly, our smartwatch app, MedSensor, performs preliminary data annotation at the edge, that is, the data origination point, which makes further downstream automation easier.

The Medhere study [39] addressed poor medication adherence by using a smartwatch accelerometer and gyroscope sensors to monitor a series of actions. While the study also considers the medication activity as a complex activity composed of atomic activities, it applies random forest ML algorithm to classify several discrete actions, including medication intake, with an average precision and recall across all activities of 0.91 and
0.93, respectively. Our study implemented an LSTM network algorithm with a similar average precision and recall of 0.91 across all activities. The LSTM injects the benefit of context as well as the architecture that suits the time series of human activity data. The Medhere study engaged 5 participants; in contrast, MedSensor engaged a larger diverse group of 28 participants.

A separate study used the Kinect depth camera to automatically generate templates for signal matching during the training phase of an inertial sensor [67]. Only 2 actions (“twist-cap” and “hand-to-mouth”) were tracked by the inertial sensor to identify the individual pill-taking signatures among the 5 participants. The approach required a Kinect depth camera during training, and the training was user specific. Despite the generally accurate prediction scores, user-specific training does not lend itself to practical scalability or generalizability across users.

Other notable medication adherence approaches include the following: (1) the AdhereTech platform focuses on a stand-alone cellular enabled bottle that transmits data in real time when it is opened and incorporates SMS text messaging, phone calls, and chime alerts and reminders to patients so that they do not forget to take their medicine; (2) the Vitality GlowCap system works with a smart cap that attaches to a medication bottle and sends alerts to patients when to take their next dose; and (3) PillsyCap is a smart-pill cap for prescription drugs [68]. The PillsyCap uses Bluetooth to synchronize with a mobile app. Although these approaches make it possible to adhere to medication requirements, human activity recognition relies on objects that are not worn on the human body. The reusability of objects, such as smart caps, is another drawback because each bottle must have its own cap. The cost of production of such objects is ultimately higher than that of relying on wearable sensor devices, such as smartwatches. In addition, smartwatches have multiple capabilities that make them appealing as affordable and useful devices.

**Limitations**

The limitation of our approach, given the current state of the technology, is the availability of data from a single hand. Therefore, activities such as smoking that can be completed by a single hand, may not be detected by a watch that is worn on the opposite hand. Fortunately, because activities such as medication taking are difficult to complete with a single hand, a residual portion of the activity will always be present from the perspective of a single watch. This problem can be easily overcome with the presence of a sensing device on each hand. Although not common presently, the arrival of smart wristbands, rings, and other forms of wearable devices is likely to provide a more complete picture of a person’s daily activities [27,30,33,35,69].

The second limitation is the method by which people may wear their watches. A watch can be worn in 4 distinct ways on the left or right hand and inside or outside of the wrist. In this study, we asked participants to wear their watch on their right hand and on the outside of their wrists. However, sensor data recorded by the same watch in any of the other 3 configurations will produce related but undistinguishable signals by the ANN. Consistently wearing a watch on the outside right-hand side is critical at the current stage of our scientific development. However, using the existing human body symmetry and the relationship between the inside and outside of the wrists, mechanisms of unifying sensor signals collected from any mode of use can be developed as demonstrated previously [20]. This will produce high-capacity models with broader experience to recognize medication gestures regardless of the watch-face orientation.

**Future Implications**

Our current protocol performed well. However, our future investigations will benefit from 2 additional steps to our existing protocol. The first step will collect calibration data during the initial orientation session. Currently, our orientation consists of familiarizing the participant with the watch, app, and use of the app. In the future, we will incorporate a second step, which will use this orientation session to collect data from a set of simple activities (eg, touch toes, touch hips, and touch head) with the watch worn on each hand to obtain useful participant-specific data at baseline. By collecting data from left and right hands, we can establish a more precise relationship between the right and left hands for the given participants. Although perfect human symmetry may indicate a 180° rotation between the 2, natural human posture may create a departure from the ideal 180° symmetry. This information can be used to allow the user to wear the watch in any preferred method and provide the necessary information for the correction that is needed for the existing right-handed ANN. We intend to add this step to our protocol and conduct a future study with participants who are taking their own medications in their natural environment.

In addition, as an ultimate objective, we aim to develop an application that can decipher numerous human activities to establish correlative or causative relationships between activities. For instance, medication-taking activity may occur at 8 PM before sleep activity or 7 AM before breakfast eating activity, or eating activity at 1 PM may subsequently be followed by cigarette smoking. The ability to monitor the temporal relationship between these events would be useful to provide the much-needed context to further understand human behavior and, therefore, to model useful health-related solutions or provide real-time intervention reminders. To accomplish this, we need to engage in a formal investigation of the optimal viewable window size for an LSTM that is sufficient to successfully decipher all activities of interest. In addition, there exists some inherent parallelism between human activities and the principles of written language. To fully leverage this parallel analogy, human activities need to be examined in the more fundamental fashion by decomposing complex activities further into their most basic building blocks, referred to in this study as mini-gestures. Our previous work [15,48] illustrates the mini-gesture decomposition of the eating activity in relation to other similar activities such as smoking. Furthermore, our study also compared the medication activity against the jogging activity, and our models confirmed little confusion between medication taking and jogging. This could be explained by the fact that the mini-activities of both complex dynamic activities are largely different. Based on this observation, we speculate the accuracy of the system to increase notably if other natural daily activities are included in our training and testing sets.
because of their dissimilarity to medication-taking gestures. Thus, our future goal is to develop a powerful and useful application that can distinguish between a multitude of different activities by recognizing the combined mini-gestures or mini-activities of more complex behaviors and temporal relationships.

Conclusions
Medication adherence is a complex human behavior that is associated with chronic condition self-management. It remains a global public health challenge, as nearly 50% of people fail to adhere to their medication regimens. Automated detection of medication-taking activities is of critical importance for improving treatment effectiveness. The automatic detection of medication-taking gestures will also help eliminate the burden of self-reporting from the participants and provide a simpler method of tracking MTEs. In this study, we demonstrated the use of LSTM to detect and recognize mini-activities and complex activities. We have demonstrated successful identification of individual medication gestures with an accuracy of approximately 93.6% when tested against activities that substantially resemble medication taking, such as smoking and eating, which share the common mini-gestures of hand-to-mouth, hand-on-mouth, and hand-off-mouth. Furthermore, we demonstrated the ability of the neural network to distinguish MTEs from the similar activities of eating and smoking and from the dissimilar activity of jogging. Our future work will build on these successes by making small changes to the protocol and further tuning the network hyperparameter values. In the short term, we anticipate testing MedSensor with people who are taking medications in their natural environment, and our long-term goal is to develop a comprehensive app that can accurately identify a multitude of human behaviors.

Acknowledgments
Funding for this work was provided by the National Institute of General Medical Sciences (grant P20 RR-01646100), College of Computer Science and Engineering, and the Advancing Chronic Care Outcomes through Research and Innovation Center, College of Nursing, University of South Carolina.

Conflicts of Interest
None declared.

References
6. Changizi M, Kaveh MH. Effectiveness of the mHealth technology in improvement of healthy behaviors in an elderly population-a systematic review. Mhealth 2017 Nov 27;3:51 [FREE Full text] [doi: 10.21037/mhealth.2017.08.06] [Medline: 29430455]
smartwatch accelerometer sensor data. arXiv 2021 May 19.
citeserv.ist.psu.edu/document?repid=rep1&type=pdf&doi=5b0a37198593fbf91aa3a613f64461c9f68be3b [accessed
2022-09-08]
on Mobile computing and networking. 1998 Presented at: MobiCom98: 4th Annual ACM International Conference on
[doi: 10.1109/98.626982]
smartwatch technology: validation study. JMIR mHealth UHealth 2017 Dec 13;5(12):e189 [FREE Full text] [doi: 10.2196/mhealth.9035]
22. Cole CA, Power S, Tomko RL, Froeliger B, Valafar H. Quantification of smoking characteristics using smartwatch
technology: pilot feasibility study of new technology. JMIR Form Res 2021 Feb 05;5(2):e20464 [FREE Full text] [doi: 10.2196/20464]
23. Imtiaz MH, Ramos-Garcia RI, Wattal S, Tiffany S, Sazonov E. Wearable sensors for monitoring of cigarette smoking in
26. Guiry J, van de Ven P, Nelson J. Multi-sensor fusion for enhanced contextual awareness of everyday activities with
IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops). 2015
Presented at: IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom
Workshops); Mar 23 - 27, 2015; St. Louis, MO, USA. [doi: 10.1109/percomw.2015.7134103]
30. Moschetti A, Fiorini L, Esposito D, Dario P, Cavallo F. Recognition of daily gestures with wearable inertial rings and
31. Ramos-Garcia RI, Muth ER, Gowdy JN, Hoover AW. Improving the recognition of eating gestures using intergesture
32. Aguilera AA, Brena RF, Mayora O, Molino-Minero-Re E, Trejo LA. Multi-sensor fusion for activity recognition-a survey.
34. Chaccour K, Darazi R, El Hassani AH, Andres E. From fall detection to fall prevention: a generic classification of fall-related
35. Qu J, Wu C, Li Q, Wang T, Soliman A. Human fall detection algorithm design based on sensor fusion and multi-threshold
36. Sun X, Qiu L, Wu Y, Tang Y, Cao G. SleepMonitor: monitoring respiratory rate and body position during sleep using
37. Peake JM, Kerr G, Sullivan JP. A critical review of consumer wearables, mobile applications, and equipment for providing
biofeedback, monitoring sleep, and sleep in physically active populations. Front Physiol 2018 Jun 28:9:743 [FREE Full
stage classification. In: Proceedings of the 2021 International Conference on Computational Science and Computational


50. Wireless sensor data mining. WISDM Lab. URL: https://www.cis.fordham.edu/wisdm/ [accessed 2022-09-14]


55. Attribution-ShareAlike 4.0 International (CC BY-SA 4.0). Creative Commons. URL: https://creativecommons.org/licenses/by-sa/4.0/ [accessed 2023-04-23]


Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI</td>
<td>artificial intelligence</td>
</tr>
<tr>
<td>ANN</td>
<td>artificial neural network</td>
</tr>
<tr>
<td>CDC</td>
<td>Centers for Disease Control and Prevention</td>
</tr>
<tr>
<td>FN</td>
<td>false negative</td>
</tr>
<tr>
<td>FP</td>
<td>false positive</td>
</tr>
<tr>
<td>LSTM</td>
<td>long short-term memory</td>
</tr>
<tr>
<td>ML</td>
<td>machine learning</td>
</tr>
<tr>
<td>MTE</td>
<td>medication-taking event</td>
</tr>
<tr>
<td>nMTE</td>
<td>natural medication-taking event</td>
</tr>
<tr>
<td>RNN</td>
<td>recurrent neural network</td>
</tr>
<tr>
<td>sMTE</td>
<td>scripted medication-taking event</td>
</tr>
<tr>
<td>TN</td>
<td>true negative</td>
</tr>
<tr>
<td>TP</td>
<td>true positive</td>
</tr>
</tbody>
</table>

Edited by A Kushniruk; submitted 14.09.22; peer-reviewed by M Kapsetaki, Z Li, Y Wang; comments to author 16.11.22; revised version received 10.01.23; accepted 11.02.23; published 04.05.23.

Please cite as:
Odhiambo CO, Ablonczy I, Wright PJ, Corbett CF, Reichardt S, Valafar H
Detecting Medication-Taking Gestures Using Machine Learning and Accelerometer Data Collected via Smartwatch Technology: Instrument Validation Study
JMIR Hum Factors 2023;10:e42714
URL: https://humanfactors.jmir.org/2023/1/e42714
doi:10.2196/42714
PMID:37140971
Behavioral Predictors of Intention to Use a Text Messaging Reminder System Among People Living With HIV in Rural Uganda: Survey Study

Jeffrey I Campbell1, MPH, MD; Isaac Aturinda2, MBA; Evans Mwesigwa2, BA; Annabella Habinka3, PhD; Michael Kanyesigye2, BSc; Richard J Holden4, PhD; Mark J Siedner5, MPH, MD; John D Kraemer6, JD, MPH

1Boston Medical Center, Boston, MA, United States
2Mbarara University of Science and Technology, Mbarara, Uganda
3Makerere University, Kampala, Uganda
4School of Public Health, Indiana University, Bloomington, IN, United States
5Massachusetts General Hospital, Massachusetts, MA, United States
6Department of Health Management and Policy, Georgetown University School of Health, Washington, DC, United States

Corresponding Author:
Jeffrey I Campbell, MPH, MD
Boston Medical Center
6th floor
670, Albany Street
Boston, MA, 02118
United States
Phone: 1 617 414 4290
Fax: 1 617 414 4290
Email: jeffreycampbell@bmc.org

Abstract

Background: The expansion of cellular phones in sub-Saharan Africa spurred the development of SMS text message–based mobile health (mHealth) technology. Numerous SMS text message–based interventions have attempted to increase retention in care for people living with HIV in sub-Saharan Africa. Many of these interventions have failed to scale. Understanding theory-grounded factors leading to mHealth acceptability is needed to create scalable, contextually appropriate, and user-focused interventions to improve longitudinal HIV care for people living with HIV in sub-Saharan Africa.

Objective: In this study, we aimed to understand the relationship between constructs from the Unified Theory of Acceptance and Use of Technology (UTAUT), constructs identified in previous qualitative research, and behavioral intention to use a novel SMS text message–based mHealth intervention designed to improve care retention among people living with HIV initiating treatment in rural Uganda.

Methods: We conducted a survey of people living with HIV who were newly initiating HIV care in Mbarara, Uganda, and had agreed to use a novel SMS text message–based system that notified them of abnormal laboratory results and reminded them to return to the clinic. Survey items assessed behavioral intention to use the SMS text messaging system; constructs from UTAUT; and demographics, literacy, SMS text messaging experience, HIV status disclosure, and social support. We used factor analysis and logistic regression to estimate the relationships between UTAUT constructs and the behavioral intention to use the SMS text messaging system.

Results: A total of 249 participants completed the surveys, of whom 115 (46.2%) expressed high behavioral intention to use the SMS text messaging intervention. In a multivariable analysis, we found that performance expectancy (adjusted odds ratio [aOR] of the scaled factor score 5.69, 95% CI 2.64-12.25; P<.001), effort expectancy (aOR of the scaled factor score 4.87, 95% CI 1.75-13.51; P=.002), and social influence (measured as a 1-unit Likert score increase in the perception that clinical staff have been helpful in the use of the SMS text messaging program; aOR 3.03, 95% CI 1.21-7.54; P=.02) were significantly associated with high behavioral intention to use the SMS text messaging program. SMS text messaging experience (aOR/1-unit increase 1.48, 95% CI 1.11-1.96; P=.008) and age (aOR/1-year increase 1.07, 95% CI 1.03-1.13; P=.003) were also significantly associated with increased odds of high intention to use the system.
Conclusions: Performance expectancy, effort expectancy, and social influence, as well as age and SMS experience, were drivers of high behavioral intention to use an SMS text messaging reminder system among people living with HIV initiating treatment in rural Uganda. These findings highlight salient factors associated with SMS intervention acceptability in this population and indicate attributes that are likely to be key to the successful development and scaling of novel mHealth interventions.

(JMIR Hum Factors 2023;10:e42952) doi:10.2196/42952

KEYWORDS
mobile health; mHealth; HIV; intention; SMS; cellular phone; cell phone; Africa; reminder; alert; notification; prompt

Introduction

Background

In 2020, there were an estimated 93 cellular phone subscriptions per 100 people living in sub-Saharan Africa [1]. Numerous interventions have attempted to leverage this widespread cellular phone coverage to address health gaps in the region, especially with respect to people living with HIV [2,3]. Retention in HIV care is complex, and barriers to long-term engagement range from individuals’ physical and mental health to resource constraints and social and cultural factors such as stigma [4]. The ubiquity of cellular phones makes mobile health (mHealth) interventions a promising tool to overcome many of these challenges to retention in care. A recent meta-analysis found that mobile phone reminders significantly improved retention in care for people living with HIV [5]. However, despite promising pilot studies, many mHealth interventions have not been scaled in sub-Saharan Africa [6,7]. Barriers such as unstable funding, unreliability of technology, and health systems’ lack of capacity to integrate electronic data can impede the broad uptake of mHealth interventions [7]. Although many studies have examined the acceptance and acceptability during intensive pilot intervention periods, there are fewer data that capture acceptability after the pilot phase.

A postpilot, theory-grounded understanding of mHealth acceptability and use among people living with HIV in sub-Saharan Africa is critical for creating sustainable, contextually adapted, user-focused interventions to improve longitudinal HIV-related care [8]. Models that integrate social and cultural contexts with perceptions of technology utility are particularly valuable to understand mHealth uptake among people living with HIV in sub-Saharan Africa, where previous studies have found that social support and influence are among the key perceived benefits of mHealth technologies [9]. Theoretical models of technology acceptance have been developed in high-resource settings to explain the links between user perceptions, social influences, intentions to use, and the actual use of technologies [10-14]. These models have also been applied to health technologies for HIV care in resource-limited settings [9,14-16]. Central to many of these models is behavioral intention, a concept that reflects prospective users’ perceived intention to use a new technology. Behavioral intention is often used as a measure of acceptance in technology acceptance models because it can be more readily measured than the actual use of many technologies and generally correlates with technology use.

In a previous qualitative study of people living with HIV in rural Uganda, we examined user attitudes toward an SMS text message notification sent to alert patients with low CD4 counts and to recall them to the clinic to initiate antiretroviral therapy (ART). On the basis of this investigation, we developed the Technology Acceptance Model for Resource-Limited Settings [17]. This model situates intended health outcomes within a framework of the behavioral factors leading to technology use and the downstream intervening factors that attenuate or propel the link between technology use and health outcomes. The qualitative study provided insights into the links between technology acceptance and anticipated health outcomes. However, as a qualitative study, it did not estimate the relationships between established technology acceptance constructs and behavioral intention to use an mHealth technology.

Objective

Here, we attempted to test the above conceptual framework by examining factors from technology acceptance theory—specifically, from the Unified Theory of Acceptance and Use of Technology (UTAUT) [13], which we hypothesized affects the behavioral intention to use SMS text message–based health interventions among people living with HIV. Our primary objective was to quantitatively estimate the relationships between these behavioral constructs and behavioral intention to use an mHealth application among intended end users, specifically people living with HIV in a sub-Saharan African setting.

Methods

Population and Setting

We conducted a standardized survey of people living with HIV who initiated HIV care at the Mbarara Regional Referral Hospital Immune Suppression Syndromes Clinic in Mbarara, Uganda. Our study was conducted in conjunction with a clinic-wide rollout of an SMS text message–based reminder intervention. On the basis of the results from a prospective before-and-after clinical trial at the same clinic (NCT01579214) [18], the intervention aimed to improve patient-provider communication, remind individuals of upcoming appointments, and notify them of laboratory results (ie, CD4 count and viral load results). ART-naive people living with HIV who were >18 years old and were initiating care at the clinic were offered voluntary enrollment in the SMS text messaging reminder program by clinic staff on the day of ART initiation. In this program, standardized SMS text messages were automatically sent 7 days and 1 day before the scheduled clinic return dates. The messages were sent in the morning. Participants could choose to receive messages in the region’s most common

https://humanfactors.jmir.org/2023/1/e42952
languages: English, Runyankole, or Luganda. The messaging application was developed by a local technology company (iStreams). It operated with a modem that interacted with the clinic’s server to send messages through a phone network (Airtel). The SMS text message contained a reminder instructing patients to return to the clinic for their appointments. On the basis of the results of a previous trial that showed decreased SMS text message uptake owing to encoded or password-protected messages [19], messages were not encoded or password protected. To avoid loss of confidentiality or disclosure, the messages did not contain patient names, nor mentioned HIV or AIDS.

Survey Design and Variable Selection
On the day of their clinic intake, we invited the first 2 to 3 patients each day who were enrolled in the SMS text messaging reminder program to participate in this survey. After providing informed consent, participants were invited to complete a detailed survey. The survey could be completed either as a self-administered written questionnaire or as an interviewer-administered verbal questionnaire per participants’ preference. The survey was available in Runyankole, the predominant language in Mbarara. A trained research assistant administered the surveys on the day of clinic enrollment. The survey contained questions about demographics (sex, age, location of housing, literacy, and education), cell phone use, HIV-related stigma, HIV disclosure, available social support, and survey measures of constructs from the UTAUT model [13]. Survey items measuring UTAUT constructs have been validated in other contexts [13] and were adapted for our study (Table 1). Notably, although self-efficacy, attitudes, and anxiety are not included in the original UTAUT model [13], we measured these constructs in our setting, given our application of this model in a new context, as has been used in other “extended” UTAUT models for health services [20]. All survey items were measured on a 5-point Likert scale (ranging from “strongly disagree” to “strongly agree”). Of note, because the survey was obtained at the time of SMS text messaging program enrollment, we sought to measure the perceived “acceptability” of the program, rather than use-based “acceptance.”

On the basis of previous qualitative research on SMS text messaging acceptability in this context [17], we measured the following constructs:

1. Demographics and HIV disease status
Participants reported their age, gender, and location of their homes during the survey. When available, the CD4 count at the time of study enrollment was retrieved from clinical records.

2. Literacy and educational attainment
To assess literacy, we asked the participants to read a short sentence in Runyankole or English according to their preferences. Participants were deemed literate if they could read all or parts of the sentence in their preferred language. Participants were asked about the highest level of education they had completed.

3. SMS text message experience
Because the ability to send an SMS text message was believed to represent cell phone familiarity, we defined texting experience as an ordered categorical variable based on the number of reported SMS text messages sent per week. We inspected a Lowess plot and found that the relationship between texting exposure and behavioral intention was logit-linear across categories. We therefore included this variable as a continuous variable in models.

4. HIV status disclosure
Previous HIV status disclosure was measured using a single binary variable that represented whether the participants had disclosed their HIV diagnosis to anyone.

5. Social support
We measured social support using a validated social support scale [21]. This scale uses 10 Likert-scale questions about family and community social support, which are averaged and dichotomized into “high social support” and “low social support” categories.
<table>
<thead>
<tr>
<th>English</th>
<th>Runyankole</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PE</strong>—degree to which the SMS system will help or be useful for patients to receive care</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PE1. I would find the SMS program useful.</td>
<td>Nka nshangire enkora y’okusindika obutumwa bwesimu obuhandikirwe eri ey’omugasho.</td>
<td><em>b</em></td>
</tr>
<tr>
<td>PE2. Using the SMS program enables me to get care from the clinic more quickly.</td>
<td>Okukozesa enkora y’okusindika obutumwa bwesimu obuhandikirwe nikimpweera kurahuka kutunga obujanjabi aha kirinika.</td>
<td>—</td>
</tr>
<tr>
<td>PE3. Using the SMS program increases my ability to get help from the clinic.</td>
<td>Okukozesa enkora y’okusindika obutumwa bwesimu obuhandikirwe nikwongyera aha kubaasakutunga obujanjabi aha kirinika.</td>
<td>—</td>
</tr>
<tr>
<td>PE4. If I use the SMS program, I will increase my chances of getting help at the clinic.</td>
<td>Nabankorise enkora y’okusindika obutumwa bwesimu obuhandikirwe, ninyija kwongyera aha migisha yangye eyoku-tunga obujanjabi aha kirinika.</td>
<td>Dropped owing to undefined estimate in factor analysis (high correlation with item PE3).</td>
</tr>
<tr>
<td><strong>EE</strong>—degree to which the SMS system is easy to use</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EE1. My interaction with the SMS program would be clear and understandable.</td>
<td>Okukoresa enkora y’okusindika obutumwa bwesimu obuhandikirwe kukabaire kurikushoboroka kandi kurikwetegerezibwa.</td>
<td>—</td>
</tr>
<tr>
<td>EE2. It would be easy for me to become skillful at using the SMS program.</td>
<td>Kikanyoroobeire okutunga obukugugu omukukoresa enkora y’okusindika obutumwa bwesimu obuhandikirwe.</td>
<td>—</td>
</tr>
<tr>
<td>EE3. I would find the SMS program easy to use.</td>
<td>Nka nshangire enkora y’okusindika obutumwa bwesimu obuhandikirwe eyoroobi kukoresa.</td>
<td>—</td>
</tr>
<tr>
<td>EE4. Learning to operate the SMS program is easy for me.</td>
<td>Okwega kukoresa enkora y’okusindika obutumwa bwesimu obuhandikirwe nikunyoroobera.</td>
<td>—</td>
</tr>
<tr>
<td><strong>ATT</strong>—degree to which patients hold positive or negative perception of using the SMS program</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATT1. Using the SMS program is a good idea.</td>
<td>Nekitekateko kirungi okukoresa enkora y’okusindika obutumwa bwesimu obuhandikirwe.</td>
<td>—</td>
</tr>
<tr>
<td>ATT2. The SMS program makes work more interesting.</td>
<td>Enkora y’okusindika obutumwa bwesimu obuhandikirwe nkyongyera omwete kwokukora.</td>
<td>—</td>
</tr>
<tr>
<td>ATT3. Working with the SMS program is fun.</td>
<td>Okukoresa enkora y’okusindika obutumwa bwesimu obuhandikirwe na mashemererwa gonka.</td>
<td>—</td>
</tr>
<tr>
<td>ATT4. I like working with the SMS program.</td>
<td>Ninkunda kukoresa enkora y’okusindika obutumwa bwesimu obuhandikirwe.</td>
<td>—</td>
</tr>
<tr>
<td><strong>SI</strong>—degree to which patients perceive that other people important to them want them to use the SMS system</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI1. People who influence my behavior think that I should use the SMS program.</td>
<td>Abantu abarikuretera natwaza nokokondikutwaza omumicwe nibatekateka ngu nshemereire kukoresa enkora y’okusindika obutumwa bwesiimu obuhandikirwe.</td>
<td>Dropped owing to undefined estimate in factor analysis (high correlation with item SI2).</td>
</tr>
<tr>
<td>SI2. People who are important to me think that I should use the SMS program.</td>
<td>Abantu abunkutwara nkab’omugasho aharinye nibatekateka ngu nshemereire kukoresa enkora y’okusindika obutumwa bwesiimu obuhandikirwe.</td>
<td>Used as a measured variable in multivariable analysis.</td>
</tr>
<tr>
<td>SI3. The clinic staff have been helpful in the use of the SMS program.</td>
<td>Abakozi ba kirinika babeire bari abahwezi omunukokoresa enkora y’okusindika obutumwa bwesimu obuhandikirwe.</td>
<td>Used as a measured variable in multivariable analysis.</td>
</tr>
<tr>
<td>SI4. In general, the clinic has supported the use of the SMS program.</td>
<td>Okutwariza hamwe, kirinika ehagiire enkora y’okusindika obutumwa bwesimu obuhandikirwe.</td>
<td>Dropped owing to undefined estimate in factor analysis (high correlation with items SI2 and SI3).</td>
</tr>
<tr>
<td><strong>FC</strong>—degree to which patients perceive sufficient resources and infrastructure to use the SMS system</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FC1. I have the resources necessary to use the SMS program.</td>
<td>Nyiine ebirikwetagisa kukoresa enkora y’okusindika obutumwa bwesiimu obuhandikirwe.</td>
<td>—</td>
</tr>
<tr>
<td>FC2. I have the knowledge necessary to use the SMS program.</td>
<td>Nyiine amagezi agarikwetagisa kukoresa enkora y’okusindika obutumwa bwesiimu obuhandikirwe.</td>
<td>—</td>
</tr>
<tr>
<td>FC3. The SMS program is not compatible with other SMS programs I use.</td>
<td>Enkora y’okusindika obutumwa bwesiimu obuhandikirwe terikukwatra nezindi nkora nekoze ezindikukoresa.</td>
<td>—</td>
</tr>
</tbody>
</table>
FC4. A specific person (or group) is available for assistance with SMS program difficulties.

Omuntu nari abantu batoraine bariho kumpwera ebizibu byenkora y’okusindika obutumwa bwesiimu obuhandikirwe. Dropped owing to poor loading in EFA.

SEb—degree to which patients perceive they have the aptitude to use the SMS system

I could successfully get the information I need using the SMS program...

Nimbaasa kutunga amakuru goona agindikwenda ndikwejunisa enkora y’okusindika obutumwa bwesiimu obuhandikirwe...

SE1. If there was no one around to tell me what to do as I go.

Habahatariho muntu wena kungambira okundatwaze nenkora egi.

SE2. If I could call someone for help if I got stuck.

Kunakuba nimbaasa kugira omuntu owunayeeta naheza kuremererwa.

SE3. If I had a lot of time, I could get the information I needed from the SMS program.

Naba nyine obwire bwingi mbaasa kutunga amakuru agindikwenda onu nkora y’okusindika obutumwa bwesiimu obuhandikirwe.

SE4. If I had just the help of the information, I received at the clinic about the SMS program.

Kurinintunga obuyambi bwekyokukora nkobunatungire aha kirinika obukwatiraine nenkora y’okusindika obutumwa bwesiimu obuhandikirwe.

Anxiety—degree to which patients feel apprehension or fear about using the SMS program

Anxiety 1. I feel apprehensive about using the SMS program.

Nyine obutagubwagye hamwe nobwooba bwokukoresa enkora y’okusindika obutumwa bwesiimu obuhandikirwe.

Anxiety 2. It scares me to think that I could lose a lot of information using the SMS program by hitting the wrong key.

Nikindetera obwooba okutekateka ngu kunakunyiiga ejepsha erigwaire mbaasa kuburwaho amakuru maangi naba ninkukoresa enkora y’okusindika obutumwa bwesiimu obuhandikirwe.

Anxiety 3. I hesitate to use the SMS program for fear of making mistakes I cannot correct.

Tinkurahukiriza kukoresa enkora y’okusindika obutumwa bwesiimu obuhandikirwe aha bwokutiina kukora enshobi enzwintarikubaasa kugoroora.

Anxiety 4. The SMS program is somewhat intimidating to me.

Enkora y’okusindika obutumwa bwesiimu obuhandikirwe nentinisamu kakye.

BIji—degree to which patients intend to use the SMS system

B1i. I intend to use the SMS program in the next 3 months.

Ninyenda kukoresa enkora y’okusindika obutumwa bwesiimu obuhandikirwe omumyezi eshatu erikwaja.

B2i. I predict I would use the SMS program in the next 3 months.

Nintebereza kwija kukoresa enkora y’okusindika obutumwa bwesiimu obuhandikirwe omumyezi eshatu erikwaja.

B3i. I plan to use the SMS program in the next 3 months.

Nintekateka kwija kukoresa enkora y’okusindika obutumwa bwesiimu obuhandikirwe omumyezi eshatu erikwaja.

Statistical Analysis

We used descriptive and summary statistics to characterize the study population. For constructs drawn from the UTAUT model, we performed exploratory factor analysis (EFA) to examine the relationship between the measured survey items and the latent constructs they measured. We elected to perform EFA (rather than confirmatory factor analysis) because we posited adaptations to the overall UTAUT model (eg, adapting questions to our setting; Table 1) and because UTAUT was designed to describe the acceptability of nonhealth technology in well-resourced settings, rather than mHealth in a sub-Saharan African context. We removed items that performed poorly in the EFAs or that yielded undefined estimates in the factor analysis, owing to high correlation with one another (Table 1). We then attempted to perform structural equation modeling to understand the relationship between behavioral intention,
UTAUT variables, and other covariates. However, the structural equation models could not be identified based on the number of latent variables, their associations, and the number of available observations. For our primary analysis, we therefore fit logistic regression models in which we represented latent predictor constructs with factor scores derived from separate measurement models for each.

Our primary outcome of interest was behavioral intention to use the SMS text messaging program. Behavioral intention was measured using 3 Likert-scale questions (Table 1). Owing to the low variability in responses and near-perfect concordance between the items, we created a binary outcome variable to characterize “high” behavioral intention (defined as listing “strongly agree” for at least 1 of the 3 behavioral intention questions) and “low” behavioral intention (defined as listing a rating of less than “strongly agree” for all the behavioral intention questions).

Our primary predictors of interest were constructs from the UTAUT model, including (1) performance expectancy, (2) effort expectancy, (3) attitudes toward technology, (4) facilitating conditions, (5) anxiety about technology use, (6) self-efficacy, and (7) social influence. We represented these constructs in the models as factor scores. To create factor scores, we first performed EFA to identify survey items with goemin-rotated loadings of >0.4 [22] for all UTAUT predictor constructs, except for social influence, which was represented as 2 observed variables. The identified survey items were then included in the confirmatory factor analysis to generate factor scores. Factor scores were divided by their IQRs so that a 1-unit increase in the score was equivalent to the difference between the middle of the bottom half and the middle of the top half. Because items measuring social influence were highly correlated, we represented this construct as 2 measured variables ("People who are important to me think that I should use the SMS program" and "The clinic staff have been helpful in the use of the SMS program") instead of as factor scores in the models. These questions were selected because they were found to have the highest face validity among questions measuring social influence. In addition, we removed one item ("If I use the SMS program, I will increase my chances of getting help at the clinic") from the measurement of performance expectancy owing to its high correlation with another performance expectancy item in the confirmatory factor analysis.

We first constructed a univariable logistic regression to evaluate the associations between UTAUT constructs and covariates with behavioral intention. We then conducted a multivariable logistic regression to identify the correlates of high (vs low) behavioral intention to use the SMS text messaging system. Owing to a priori hypotheses about their effect on behavioral intention, all covariates were included in the multivariable model. Owing to collinearity between the UTAUT constructs’ factor scores, we used a forward selection strategy to identify UTAUT constructs that were independent significant predictors of high versus low behavioral intention, using \( P < 0.20 \) for inclusion in the final model. The significance in the final model was set at \( P < 0.05 \). Patients with missing survey data (n=3) were excluded from the multivariable analysis. Patients were only excluded from specific univariable tests and factor analysis when the items considered in the specific test or analysis were missing. Because odds ratios are liable to be misinterpreted when the outcome of interest is not rare, we calculated and graphed associations as adjusted differences in the probability of the outcome following regression. We did this by using the average marginal effects with other covariates held at the observed levels.

We conducted 3 sensitivity analyses with multivariable logistic regression models using low versus high behavioral intention as the outcome and different variable inclusion strategies to select UTAUT constructs and covariates. First, we constructed a multivariable model that included all UTAUT constructs, with age and sex as covariates. Second, we created a model that included all UTAUT constructs and all covariates. Finally, we constructed models that included each UTAUT construct separately plus all covariates.

Analyses were conducted using Stata (version 17; StataCorp) and Mplus (version 8; Muthén & Muthén), which was run within Stata via the runmplus suite of commands.

**Ethics Approval**

The study was reviewed and approved by the Ethics Review Committee of the Mbarara University of Science and Technology (13/10-15), the Institutional Review Board of Massachusetts General Hospital (2015P002572), and the Uganda National Council for Science and Technology (SS 4008).

**Results**

**Participant Characteristics**

A total of 249 participants were enrolled in the survey study and completed the surveys. The mean age was 30.6 (SD 9.2) years, and 56.2% (140/249) of the patients were female. The median CD4 count at the time of enrollment was 311 (IQR 145-524). Of the 249 participants, 226 (90.8%) participants were literate and 219 (88%) endorsed high social support. Only 57.4% (143/249) of the participants noted that they had disclosed their HIV status. The majority (155/249, 62.2%) had sent fewer than 3 SMS text messages during the preceding week. Multimedia Appendix 1 summarizes the factor loadings for the survey items measuring the UTAUT constructs.

**Predictors of High (vs Low) Behavioral Intention**

In the univariable analysis, literacy, number of SMS text messages sent in the preceding week, and social support were significantly associated with intention to use the SMS text messaging program (Table 2). The mean scaled factor scores and measured survey responses (for the 2 social influence variables) differed significantly between participants with low and high behavioral intention (Table 2).
| Table 2. Univariable and multivariable analysis of predictors of behavioral intention. |
|-------------------------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                                                  | Low Intention   | High Intention  | Univariable analysis | Multivariable analysis |
|                                                  | (n=134)         | (n=115)         | P value | OR\(^a\) (95% CI) | P value | aOR\(^b\) (95% CI)² | P value |
| Patient characteristics                          |                 |                 |                     |                     |                     |                     |                     |
| Age (years), mean (SD)                           | 29.8 (8.6)      | 31.6 (9.8)      | 1.02 (0.99-1.05)    | .12                | 1.07 (1.03-1.13)    | .003 \(^d\)        |
| Sex, n (%)                                       |                 |                 |                     |                     |                     |                     |                     |
| Male                                             | 55 (41)         | 54 (47)         | Ref\(^e\)           | Ref                | Ref               | Ref                |
| Female                                           | 79 (59)         | 61 (53)         | 0.79 (0.46-1.30)    | .35                | 1.62 (0.73-3.57)    | .24                |
| Literate, n (%)                                  |                 |                 |                     |                     |                     |                     |                     |
| No                                               | 19 (14)         | 4 (3)           | Ref\(^f\)           | Ref                | Ref               | Ref                |
| Yes                                              | 115 (86)        | 111 (97)        | 4.58 (1.51-13.90)   | .007               | 0.18 (0.03-1.18)    | .07                |
| SMS sent in the past week (per 1-unit increase in the ordinal scale), n (%) |               |                 |                     |                     |                     |                     |                     |
| None                                             | 89 (67)         | 35 (30)         |                     |                    |                   |                    |
| 1-2                                              | 16 (12)         | 15 (13)         |                     |                    |                   |                    |
| 3-5                                              | 12 (9)          | 16 (14)         |                     |                    |                   |                    |
| 6-10                                             | 12 (9)          | 23 (20)         |                     |                    |                   |                    |
| 11-20                                            | 2 (2)           | 19 (17)         |                     |                    |                   |                    |
| >20                                              | 2 (2)           | 7 (6)           |                     |                    |                   |                    |
| High social support, n (%)                       |                 |                 |                     |                     |                     |                     |                     |
| No                                               | 22 (16)         | 8 (7)           | Ref\(^g\)           | Ref                | Ref               | Ref                |
| Yes                                              | 112 (84)        | 107 (93)        | 2.62 (1.12-6.16)    | .03                | 1.06 (0.33-3.44)    | .92                |
| History of prior disclosure, n (%)               |                 |                 |                     |                     |                     |                     |                     |
| No                                               | 51 (38)         | 55 (48)         | Ref\(^h\)           | Ref                | Ref               | Ref                |
| Yes                                              | 83 (62)         | 60 (52)         | 0.67 (0.40-1.11)    | .12                | 0.47 (0.21-1.02)    | .06                |
| UTAUT\(^i\) constructs                          |                 |                 |                     |                     |                     |                     |                     |
| Performance expectancy, mean scaled factor score  | −0.41           | 0.51            | 11.60 (6.39-21.03)\(^g\) | <.001            | 5.69 (2.64-12.25)\(^g\) | <.001            |
| Effort expectancy, mean scaled factor score       | −0.47           | 0.56            | 12.32 (6.37-23.81)\(^g\) | <.001            | 4.87 (1.75-13.51)\(^g\) | .002            |
| Attitudes, mean scaled factor score              | −0.27           | 0.31            | 9.93 (5.45-18.11)\(^g\) | <.001            | N/A\(^h\)         | N/A            |
| Facilitating conditions, mean scaled factor score | −0.22           | 0.25            | 10.77 (5.72-20.28)\(^g\) | <.001            | N/A\(^h\)         | N/A            |
| Anxiety, mean scaled factor score                | 0.33            | −0.38           | 0.23 (0.13-0.35)\(^g\) | <.001            | N/A\(^h\)         | N/A            |
| Self-efficacy, mean scaled factor score          | −0.42           | 0.51            | 7.61 (4.37-13.26)\(^g\) | <.001            | N/A\(^h\)         | N/A            |
| Social influence (“The clinic staff have been helpful in the use of the SMS program”), mean Likert response | 3.11            | 3.34            | 3.59 (1.92-6.71)\(^d\) | <.001            | 3.03 (1.21-7.54)\(^d\) | .02            |
| Social influence (“People who are important to me think that I should use the SMS program”), mean Likert response | 2.08            | 2.83            | 2.60 (1.87-3.61)\(^d\) | <.001            | N/A\(^h\) | N/A            |

\(^a\)OR: odds ratio.
Adjusted for age, sex, literacy, SMS experience, social support, disclosure status, performance expectancy (normalized factor score), effort expectancy (normalized factor score), and social influence (Likert-scale question).

Values in italics represent significant $P$ value.

Ref: reference.


Odds ratio associated with a 1-unit change in the factor score interquartile interval.

N/A: not applicable; excluded from the multivariable model during forward variable selection.

Odds ratio associated with a 1-unit change in the Likert scale.

A total of 1.2% (3/249) of the patients were excluded from the multivariable analysis owing to missing data. In the multivariable analysis, we found that performance expectancy (adjusted odds ratio [aOR] 5.69, 95% CI 2.64-12.25; $P<.001$); effort expectancy (aOR 4.87, 95% CI 1.75-13.51; $P=.002$); and social influence (measured as the perception that clinical staff have been helpful in the use of the SMS text messaging program; aOR 3.03, 95% CI 1.21-7.54; $P=.02$) were significantly associated with a high behavioral intention to use the SMS text messaging program (Figure 1; Table 2). Texting experience (aOR/1-unit increase 1.48, 95% CI 1.11-1.96; $P=.008$) and age (aOR/1-year increase 1.07, 95% CI 1.03-1.13; $P=.003$) were also significantly associated with increased odds of high intention. Multimedia Appendix 2 demonstrates the marginal effects of social influence, performance expectancy, and effort expectancy on the probability of high behavioral intention. Multimedia Appendix 3 shows the marginal effects of covariates on the probability of high behavioral intention.

Sensitivity Analyses

Sensitivity analyses using different strategies to select covariates for multivariable models yielded results similar to those of the main model. First, in a model that included UTAUT constructs and the covariates age and sex, performance expectancy (aOR 6.40, 95% CI 2.80-14.66; $P<.001$) and social influence (measured as the perception that “The clinic staff have been helpful in the use of the SMS program”; aOR 3.20, 95% CI 1.23-8.38; $P=.02$) were again significantly associated with behavioral intention (Multimedia Appendix 4). In both sensitivity models, there was evidence of collinearity between the UTAUT constructs, with some variables demonstrating inflated CIs. Finally, when each UTAUT construct was considered in models adjusting for covariates, but no other UTAUT constructs were considered, we found that each construct to be significantly associated with behavioral intention (Multimedia Appendix 4).
Discussion

Principal Findings

We found that performance expectancy, effort expectancy, and social influence were significantly associated with high behavioral intention to use an SMS text message–clinic return reminder system among people living with HIV initiating ART in rural Uganda. Although participants were surveyed at the time of texting program initiation, these findings suggest that texting programs that are perceived as useful, low effort, and socially supported or promoted are likely to be most acceptable and engaging to this target population. These results add important behavioral acceptability data to the field of mHealth technology acceptance and acceptability in the region. The significance of performance expectancy and effort expectancy in our study suggests that this study population values technology that is useful and easy to use when estimating their intention to use a new mHealth technology. In addition, the significance of social influence—specifically, a question regarding clinic staff support for use of the SMS text messaging system—suggests the importance of positive social norms, particularly positive impressions from health care team members, in motivating the use of mHealth-based services in this setting.

The UTAUT model has been used and adapted in various resource-limited settings to understand the intention to use new health technologies [23]. Our finding that performance expectancy, effort expectancy, and social influence were significantly associated with behavioral intention to use an SMS text messaging system in Uganda aligns with previous research on mHealth interventions in similar settings. A recent survey and qualitative literature on mobile phone–based health interventions in Africa illustrated the differential importance of the UTAUT constructs. A path analysis of UTAUT constructs examined attitudes toward a mobile interactive voice response system for monitoring childhood illness in Ghana and found that performance expectancy, effort expectancy, and social influence were positively associated with behavioral intention to use the technology [24]. Similarly, a real-world mixed methods acceptability study of an SMS text message–based adherence-monitoring intervention in Uganda concluded that performance expectancy was the key driver of the acceptability of the system [9]. Analogously, qualitative studies from Uganda have found that appealing aspects of mHealth interventions for long-term HIV care include their ability to overcome forgetfulness and stigma, whereas technical issues that could make these interventions more effortful to use have been highlighted as areas for improvement [25,26].

In contrast, in our study, attitudes toward SMS text messaging technology, facilitating conditions, anxiety, and self-efficacy were not significantly associated with behavioral intention to use the technology in a model that also included performance expectancy, effort expectancy, social influence, and covariates. In the original UTAUT model, attitudes, anxiety, and self-efficacy were not found to significantly influence behavioral intention in models that included effort expectancy (which subsumed anxiety, self-efficacy, and attitudes) and performance expectancy (which subsumed attitudes) [13], although these factors were included in subsequent extended UTAUT models for resource-limited settings [20]. Although significant in our models in which attitudes, anxiety, and self-efficacy were the sole UTAUT constructs, our findings suggest that these constructs do not explain behavioral intention in our setting beyond performance expectancy, effort expectancy, and social influence in our population and context. In addition, the lack of significance of attitudes, anxiety, and self-efficacy in the multivariable model could have been due to the relative simplicity of the SMS text messaging system and the penetration of SMS text messages into society (ie, 180/249, 72.3% of respondents reported knowing how to send an SMS). Finally, this finding may also be related to the fact that patients were surveyed before using the SMS text messaging system, making these questions hypothetical for the respondents. Given the challenges with scaling mHealth interventions and the field’s reliance on cross-sectional data to understand acceptability, longitudinal studies of how factors affect their actual use in practice will be valuable to better understand the realized impact of mHealth interventions. For example, perceived self-efficacy or anxiety about technology may become more salient, whereas the importance of effort expectancy may wane over time as people living with HIV become more familiar and versatile with a new technology. In related research, comfort with mHealth tools was found to increase over time among individuals with tuberculosis in high-resource settings [27].

We controlled for several participant characteristics hypothesized to affect the intention to use the SMS text messaging intervention. We found that increasing age was associated with higher behavioral intention. This finding was unexpected; we initially hypothesized that younger people living with HIV would be more comfortable with SMS text messaging technology in general and hence would be more likely to have high behavioral intention. By contrast, younger individuals may have expected more sophisticated technology than a simple SMS text message. Notably, in the pilot study preceding our study, increasing age was associated with decreased time to return to the clinic after an abnormal CD4 result, suggesting that increasing age may be independently associated with the propensity to engage in care [18]. Increased SMS text messaging experience may have offset anxiety about using the SMS text messaging intervention. Most participants in our study had relatively low exposure to SMS text messages (155/249, 62.2% had sent fewer than 3 SMS text messages per week). Our results suggest that as cell phones become more common and SMS text messaging becomes less expensive in resource-limited settings, SMS text message–based interventions may become more acceptable.

Our findings are generally consistent with our previous qualitatively derived frameworks for mHealth acceptability in resource-limited settings [17]. In that analysis, we interviewed people living with HIV who had used the SMS text messaging system to facilitate their return to the clinic. Participants identified factors affecting actual technology use and downstream mediators of the target health outcome (return to the clinic). Perceived ease of use and perceived usefulness were found to be upstream promoters of technology use, which was
also affected by confidentiality and disclosure considerations. In this study, the findings that performance expectancy and effort expectancy predicted high behavioral intention corroborate our previous qualitative findings that perceived ease of use and perceived usefulness promote technology use. The UTAUT constructs of performance expectancy and effort expectancy are founded in part upon the concepts of perceived usefulness and perceived ease of use, respectively, as delineated in the technology acceptance model [13], which in turn anchored our Technology Acceptance Model for Resource-Limited Settings framework. Although a history of prior disclosure was not significantly associated with high behavioral intention in our study, our survey study was unlikely to be able to capture the complex relationship between HIV-directed mHealth technology adoption and concerns about disclosure and stigma that have been captured in qualitative research from this setting [17,28].

Strengths and Limitations
The strengths of this study include the large sample size, enrollment of people living with HIV initiating ART (thus providing a unique insight into the attitudes of this important population), the use of existing technology acceptance theory, and the ability to compare this quantitative analysis with our previous qualitative research in the same setting and with a similar patient population. Our study has several limitations. First, data were obtained from a single setting in rural Uganda, and these findings may not be generalizable to other contexts in which mHealth interventions are used. Second, we adapted and translated a UTAUT questionnaire validated in high-resource settings so that it would be understandably relevant to our study population. Given the low variability in responses to behavioral intention questions, we analyzed behavioral intention as a dichotomous variable. Social desirability biases among participants may have resulted in positive responses to behavioral intention questions, which we overcame by dichotomizing “high” versus “not high” behavioral intention. In addition, the nuanced differences in the terminology used in the English behavioral intention questions may not have been translated thoroughly into Runyankole, limiting variability in responses. Given our modification of the questionnaire and the analysis of the behavioral intention variable, our findings should be considered exploratory. Third, despite our attempts to adapt survey questions to our population and setting, some survey questions did not function well in our analysis, and low variability in responses led to the inability to perform structural equation modeling. Our use of exploratory and confirmatory factor analysis was able to eliminate noncontributory questions.

Conclusions
In conclusion, we found that performance expectancy, effort expectancy, and social influence were key drivers of high behavioral intention to use an SMS text messaging reminder system among people living with HIV who had initiated ART in rural Uganda. Age and texting experience were also associated with high behavioral intention to use the texting system. These findings highlight the key ingredients of acceptable mHealth interventions in this population. Our study also suggests the need for longitudinal technology acceptability data among people living with HIV in resource-limited settings to better understand how acceptability changes over time among patients for whom long-term engagement in care is paramount.

Acknowledgments
The authors would like to thank the participants of this study and the staff of the Mbarara Regional Referral Hospital Immune Suppression Clinic. This study was funded by the National Institute of Health (K23 MH099916). JIC was supported by the Agency for Healthcare Research and Quality grant number T32 HS000063 as part of the Harvardwide Pediatric Health Services Research Fellowship Program. MJS received funding from the National Institutes of Health (K24 HL166204). The findings and conclusions in this report are those of the authors and do not necessarily represent the official views of the US National Institutes of Health.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Exploratory factor analysis loadings for items measuring Unified Theory of Acceptance and Use of Technology constructs. [DOCX File, 16 KB - humanfactors_v10i1e42952_app1.docx ]

Multimedia Appendix 2
Marginal effects of social influence, performance expectancy, and effort expectancy on the probability of high behavioral intention. [PPTX File, 153 KB - humanfactors_v10i1e42952_app2.pptx ]

Multimedia Appendix 3
Marginal effects of covariates on the probability of high behavioral intention. [PPTX File, 231 KB - humanfactors_v10i1e42952_app3.pptx ]

Multimedia Appendix 4
Results of sensitivity analyses. Association of covariates with behavioral intention, with no Unified Theory of Acceptance and Use of Technology (UTAUT) construct (column 1), and association of each UTAUT construct adjusting for covariates but not other UTAUT constructs (columns 2-9). Effect estimates and confidence intervals for covariates are shown only once, and did not vary substantially.

[DOCX File, 16 KB - humanfactors_v10i1e42952_app4.docx]

References

1. Mobile cellular subscriptions (per 100 people) - Sub-Saharan Africa. The World Bank. URL: https://data.worldbank.org/indicator/IT.CEL.SETS.P2?locations=ZG [accessed 2022-02-08]


Abbreviations
- aOR: adjusted odds ratio
- ART: antiretroviral therapy
- EFA: exploratory factor analysis
- mHealth: mobile health
- UTAUT: Unified Theory of Acceptance and Use of Technology

©Jeffrey I Campbell, Isaac Aturinda, Evans Mwesigwa, Annabella Habinka, Michael Kanyesigye, Richard J Holden, Mark J Siedner, John D Kraemer. Originally published in JIMR Human Factors (https://humanfactors.jmir.org), 05.05.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JIMR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Implementation of a Digital Health Tool for Patients Awaiting Input From a Specialist Weight Management Team: Observational Study

Petra Hanson¹,²,³, BSc, MBChB, MRCP, PhD; Charlotte Summers¹,⁴, BSc; Arjun Panesar¹,⁴, MEng; Alexandros Leonidas Liarakos²,³, MRCP, MD; Dominic Odouro-Donkor¹,²,³, MBChB, MRCP; Danniella Whyte Oshodi¹,²,³, BSc, MBChB; Luke Hailston²,³, BSc, MBBS; Harpal Randeva¹,²,³, MBChB, PhD; Vinod Menon¹,²,³, MBBS, MS; Michaela de la Fosse⁴, BSc; Amit Kaura⁵,⁶, BSc, MSc, MBChB, MRCP, PhD; Emma Shuttlewood²,³, BSc, PG Dip, PhD; Mark Loveder²,³, BSc, PhD; Donna Poole²,³, BSc, MSc, PG Dip, PhD; Thomas M Barber¹,²,³, MA, MBBS, DPhil

¹Warwick Medical School, Coventry, United Kingdom
²Warwickshire Institute for the Study of Diabetes Endocrinology and Metabolism, University Hospitals Coventry and Warwickshire National Health Services Trust, Coventry, United Kingdom
³Digital & Data Driven Research Unit, University Hospitals Coventry and Warwickshire National Health Services Trust, Coventry, United Kingdom
⁴Diabetes Digital Media Health, Coventry, United Kingdom
⁵National Institute for Health Research Imperial Biomedical Research Centre, Imperial College London and Imperial College Healthcare National Health Services Trust, London, United Kingdom
⁶National Heart & Lung Institute, Imperial College London, London, United Kingdom

Corresponding Author:
Amit Kaura, BSc, MSc, MBChB, MRCP, PhD
National Heart & Lung Institute
Imperial College London
9th Floor, Building E - Sir Michael Uren
White City Campus
London, W12 7ED
United Kingdom
Phone: 44 (0)20 7594 5735
Email: a.kaura@imperial.ac.uk

Abstract

Background: Digital tools are increasingly used on a population level as a weight loss strategy for people living with overweight and obesity. Evidence supports the feasibility of digital tools for the management of obesity in a community setting, but there is only emerging evidence for the feasibility of such tools in specialist weight management services. No study has assessed the uptake of digital tools among patients awaiting their first appointment with a specialist weight management service.

Objective: The objective of this study was to understand interest, acceptance, and engagement with a digital behavioral change platform to support specialist weight management.

Methods: This was an observational study registered as a service innovation. All patients on the waiting list for a first appointment in the tier 3 weight management service at University Hospitals Coventry and Warwickshire National Health Service (NHS) Trust were eligible to access the NHS-approved digital tool. Data on interest and engagement with the digital tool were collected. Routine clinical data were used to describe patient demographics. Focus groups were held to explore patients’ views on the use of digital tools as part of a specialist weight management service.

Results: A total of 199 patients on the waiting list were informed about the available digital tool. Just over a half (n=102, 51.3%) of patients were interested in using the app, with over one-third (n=68, 34%) of all patients engaging with the app. Overall, a third of patients on the waiting list (n=63, 32%) did not respond to the invite and 34 (17%) of patients expressed no interest in the app. Emotional eating and higher BMI was associated with interest in the Gro Health app. Male gender was associated with reduced engagement with the app. There were no differences in interest in the Gro Health app according to age, ethnicity, metabolic measures of glycemia, and lipid profile.

Conclusions: It is feasible to offer digital tools such as Gro Health to patients awaiting their first appointment with specialist weight management services. Future research should explore barriers and facilitators of engagement with digital tools. Additionally, there is a need to further evaluate the effectiveness of such tools in specialist weight management services.
weight management; precision health; digital health, hospital; secondary care; tier 3 weight management; National Health Service; weight; obese; obesity; focus group; perspective; opinion; attitude; behavior change; behavior change; mHealth; mobile health; health app

Introduction

Background

Obesity is a leading cause of chronic disease in the 21st century [1]. Despite ongoing research and innovative approaches to prevent and treat obesity, its prevalence continues to increase globally [2]. Our traditional approach to obesity management, including advice on lifestyle changes in real-world settings from health care professionals, is costly and not sustainable, given the increasing demand for health care services. The COVID-19 pandemic highlighted the importance of developing efficient strategies for weight management. With the rates of referrals to our own specialist weight management service at University Hospitals Coventry and Warwickshire (UHCW) National Health Service (NHS) Trust rising by 530% between 2014 (207 patients referred) and 2019 (1319 patients referred), waiting times for referred patients continue to increase. This exacerbates the problem as people most in need are not getting the timely support they so desperately need.

Digital tools have huge potential to transform weight management services. Current evidence shows the emerging effectiveness and weight loss potential of digital health interventions for weight loss in community settings, through the facilitation of positive behavioral changes [3,4]. The application of digital health interventions can result in up to 13% weight loss at 4 months [5] and 7.6% weight loss at 12 months [3]. Indeed, at least in the short term (less than 6 months), such interventions result in greater weight loss than more traditional face-to-face interventions [6], with apparent equivalence in the overall effectiveness between these 2 approaches in the long term (12 months) [3]. Similar findings were observed in a feasibility study of the Low Carb Program app in our obesity service at UHCW, whereby digital tool interventions for diet combined with medical appointments resulted in a similar weight loss to that from a traditionally delivered obesity service [7]. Interestingly, there are no studies that explore the feasibility of offering digital tools to patients on a waiting list for a specialist weight management service, defined as a service comprising specialist dietitians, physicians, and psychologists. This approach may provide initial support and information provision prior to engaging with the hospital obesity service, with the potential to also result in effective weight management in newly referred obese patients. A need for more evidence on this topic was highlighted by a recent meta-analysis by Berry et al [8], who highlighted the need for future studies exploring the effectiveness of digital interventions as an adjunct to specialist weight management services.

Poor uptake and engagement with digital tools remain common challenges with digital health interventions [3,6]. A progressive reduction of user engagement over time may explain a greater weight loss during the initial 6 months of use, with subsequent plateauing of body weight. Within the current literature, there are relatively few studies on how to improve and optimize user take-up and engagement with digital tools, particularly within weight management [9].

Another factor that may contribute to the poor uptake and long-term engagement with digital tools is the specificity of such apps, which generally only address 1 aspect of lifestyle—for instance, the Low Carb Program, which focuses on diet, or Strava, which focuses on exercise. In addition, support for patients from ethnic minorities is usually limited due to apps being available solely in English. Previous evidence has shown that obesity management requires a holistic, health-centered approach [10]. Lifestyle medicine has determined the 6 pillars of lifestyle to be healthy eating, physical activity, restful sleep, stress management, avoidance of risky substances such as alcohol and smoking, and healthy relationships [11]. Additionally, culturally appropriate education has shown consistent benefits over conventional care in terms of metabolic control and condition knowledge [12].

The aim of our study was to investigate the patient demographic and clinical characteristics predictive of expression of interest and subsequent engagement with a digital health weight loss tool among patients referred to tier 3 Specialist obesity service. This study was undertaken as part of a Topol Digital Fellowship funded by Health Education England.

Objective

Our primary objective was to gauge the general interest in a holistic digital health tool, Gro Health, among newly referred patients awaiting input from our obesity team and to explore the predictors of patient engagement with such digital tools. Our secondary objectives were to gain insight into how to improve the engagement of future patients referred to such digital tools within our obesity service through participant dropout rates and analysis of patient feedback on acceptability and desired features of the digital tool.

Methods

Recruitment

We offered access to the NHS-approved digital health tool Gro Health to all patients awaiting their first appointment with our hospital-based (tier 3) specialist weight management team at UHCW between January 2021 and April 2021. All eligible patients were contacted during this period by letter, phone, and email and provided with relevant details about Gro Health app.

Those patients who expressed an interest in using the tool were invited to receive usual medical care. All patients who were interested in using the app or did not respond to their initial interest in using the app were contacted during this period by letter, phone, and email and provided with relevant details about Gro Health app. Patients who were not interested in using the app or did not respond to their initial invite continued to receive usual medical care. All patients who...
were interested in using the Gro Health app also received the usual clinical care within our obesity service. Therefore, usual clinical care within the obesity service was not influenced by the patient’s interest in the Gro Health app.

Research Design
This was an observational study, registered as a service evaluation with UHCW research and development department. With clinical data extraction from routine clinical care, formal research ethics committee approval was deemed unnecessary, and no specific consent for this was necessary from patients. Patients did not receive any payment for engagement with the digital tool. Free access to the digital tool was offered to all people on the waiting list as part of the standard of care. Prior to the first use of the digital tool, each person provided informed consent to use the Gro Health app and consent for their anonymized self-reported data to be used for research purposes. No identifiable data were provided from the use of Gro Health app. Patients were invited to participate in a patient engagement workshop (lasting 1 hour) to explore their views on the use of digital tools (both generally and Gro Health app specifically) in specialist obesity services. These were held using Microsoft Teams, and participants were offered an Amazon voucher (£20; US $25) in return for their participation. For the patient engagement workshops, participants provided verbal consent to participate, record the discussion and were reminded of confidential matter of discussion at patient engagement workshops.

Intervention
Gro Health (Diabetes Digital Media) is an accessible behavior change platform that supports users to self-manage their condition and achieve their self-selected health goals through a holistic approach to health. This encompasses 4 therapeutic areas including mental well-being, sleep, activity, and nutrition. The Gro Health platform facilitates precision digital health by providing evidence-based structured education, guided behavioral change activities, weekly virtual meetups and community support, health tracking, and data-driven insights to users based on their individualized data collected on signup. The user experience is tailored to self-selected health goals, ethnicity, gender, dietary preference, and levels of activity. Gro Health uses the capability, opportunity, motivation, behavior (COM-B) model of behavior change, which identifies 3 factors that need to be present for any behavior to occur: capability, opportunity, and motivation. These factors interact over time so that behavior is seen as part of a dynamic system with positive and negative feedback loops. To create a sustainable behavior change environment and support users with diverse needs and levels of accessibility, Gro Health is offered across a variety of platforms that include web-based (responsive), iOS, Android, Apple/Google Watch, Smart TV, and digital assistants such as Google Hub and Amazon Alexa in multiple languages (English, French, German, and Hindi) to support the local population. A clinical dashboard enables the clinical team to remotely assess user engagement with the app. A recently reported study demonstrated that the users of Gro Health had improvements in symptoms of stress, anxiety, and depression measured through standardized questionnaires over 12 weeks [13]. Please see Multimedia Appendix 1 for more details on the Gro Health platform’s precision health components.

Data Collection and Statistical Analysis
Baseline patient and clinical characteristics including age, gender, weight, BMI, ethnicity, blood test results, and psychological surveys were collected as part of routine clinical care in our obesity service. Psychological data collected from patients within our obesity service in the past were used as a control for comparing psychological variables. Patients were categorized into four mutually exclusive groups based on their responses to have free access to the Gro Health app: (1) those who were interested in using the app but did not engage with it, (2) those who were interested in and engaged with the app, (3) those who were not interested in and did not engage with the app, and (4) those who did not respond to the offer of access to the app. App engagement was defined as having opened the app and imputed data within the last month (data collected in August 2021, 4 months after the last person registered with the app). Patients who were not interested in using the digital tool were able to provide a reason for this decision. Patients who registered with the app were asked to complete an anonymized feedback form, using open-ended questions, which was sent to patients via email or were invited to provide feedback and share opinion on using digital tools during patient engagement workshops held between April and May 2021. The workshops were recorded, and the main themes were summarized by the researcher who led these workshops. The hospital lead for patient and public involvement had an oversight of these workshops that complied with the UK standard for public involvement [14]. Formal qualitative analysis from patient engagement workshops was not done as this was outside of the scope of this service evaluation. Themes from workshops contributed to the development of a bespoke digital product.
Anonymous data on engagement with the app were analyzed in August 2021. Psychological data collected routinely in our service were used as a control group for comparison of psychological data of newly referred patients who were interested in using the app. Psychological surveys routinely collected consisted of these validated tools: a brief measure for generalized anxiety disorder (GAD-7) [15], a brief depression severity measure (Patient Health Questionnaire-9; PHQ-9) [16], the Warwick-Edinburgh Mental Wellbeing Scale [17], and the Dutch Eating Behavior Questionnaire [18]. A score of ≥10 is considered clinically significant on the GAD-7 and PHQ-9 questionnaires, indicating a likelihood of anxiety and depression. SPSS (version 27; IBM Corp) and R (R Development Core Team) were used to analyze data. Normal distribution was assessed with Shapiro-Wilk test. Nonparametric data were analyzed with Mann-Whitney U test. Parametric data were analyzed with an independent 2-tailed t test. A multivariable multinomial logistic regression model was used to evaluate the association between patient characteristics and user groups (interested and engaged, interested and not engaged, refused, and not responded). The reference group for analyses was the interested and engaged group.
Results

Descriptive Statistics
All patients awaiting their first appointment with the UHCW obesity team (N=199) were contacted between January and April 2021 and offered free access to the Gro Health app. Figure 1 summarizes the flowchart of study participants. Engagement with the app was assessed in August 2021.

Figure 1. Flowchart of patient inclusion.

The baseline characteristics of the cohort of patients offered the Gro Health app are summarized in Table 1. All data, except for low-density lipoprotein cholesterol, were not normally distributed.

Just over half (n=102, 51.3%) of patients were interested in using the app, with over one-third (n=68, 34.2%) of these patients engaged with the app. Of the patients who were not interested in using the app, the responses received for their rationale for this decision were categorized as shown in Table 2.

Table 1. Baseline characteristics of the whole cohort (N=199).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age range (years)</td>
<td>18-81</td>
</tr>
<tr>
<td>Age (years), median (IQR)</td>
<td>40 (32-51)</td>
</tr>
<tr>
<td>BMI (kg/m²), median (IQR)</td>
<td>45.5 (41.9-51)</td>
</tr>
<tr>
<td>Weight (kg), median (IQR)</td>
<td>130 (114.3-148)</td>
</tr>
<tr>
<td>Female, n (%)</td>
<td>154 (77.4)</td>
</tr>
</tbody>
</table>

aData on BMI were available for 193 patients.
bData on body weight were available for 167 patients.

Table 2. Reasons for declining an offer to use the Gro Health app among respondents (n=34).

<table>
<thead>
<tr>
<th>Reasons</th>
<th>Respondents, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actively involved in a research trial</td>
<td>7 (21)</td>
</tr>
<tr>
<td>Already seen by a weight management clinician</td>
<td>11 (32)</td>
</tr>
<tr>
<td>Only surgery wanted or lost weight already</td>
<td>3 (9)</td>
</tr>
<tr>
<td>No smartphone or internet</td>
<td>4 (12)</td>
</tr>
<tr>
<td>Using other apps</td>
<td>1 (3)</td>
</tr>
<tr>
<td>Not interested in apps</td>
<td>4 (12)</td>
</tr>
<tr>
<td>Other reasons (died, not happy to tell us the details)</td>
<td>4 (12)</td>
</tr>
</tbody>
</table>

Overall, a third of patients on the waiting list (n=63, 32%) did not respond to communication attempts via telephone, postal letter, or email correspondence. Table 3 summarizes the main characteristics of patients in the 4 groups.

Among patients who were interested in using the app, those not engaged were more likely to be male than those who were engaged (odds ratio 6.17, 95% CI 1.22-31.20; P=.03). There were no differences between the user groups according to age or ethnicity. Patients who did not respond were more likely to...
have a lower BMI when compared to the BMI of patients who were interested and engaged with the app (0.89 kg/m$^2$, 95% CI 0.81-1.00; $P=.05$). The results are summarized in Table 4.

Table 3. Summary of 4 user groups.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Interested and engaged (n=68)</th>
<th>Interested but not engaged (n=34)</th>
<th>Refused (n=34)</th>
<th>Did not respond (n=63)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age range (years)</td>
<td>18-71</td>
<td>19-81</td>
<td>21-76</td>
<td>19-69</td>
</tr>
<tr>
<td>Age (years), median (IQR)</td>
<td>39 (31-48)</td>
<td>47 (36.3-55)</td>
<td>45.5 (33-53)</td>
<td>38 (31-52)</td>
</tr>
<tr>
<td>BMI (kg/m$^2$), median (IQR)</td>
<td>46 (41.9-50.5)</td>
<td>45.3 (41.9-49.8)</td>
<td>45.6 (41.6-52.8)</td>
<td>45 (42-50.6)</td>
</tr>
<tr>
<td>Weight (kg), median (IQR)</td>
<td>128.3 (112.8-143)</td>
<td>130 (116-149.2)</td>
<td>130 (111.4-144.9)</td>
<td>134 (120-151.4)</td>
</tr>
</tbody>
</table>

Sex

<table>
<thead>
<tr>
<th></th>
<th>Female, n (%)</th>
<th>Male, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any Black background</td>
<td>1 (1)</td>
<td>2 (6)</td>
</tr>
<tr>
<td>White</td>
<td>46 (68)</td>
<td>19 (56)</td>
</tr>
<tr>
<td>Any Asian background</td>
<td>2 (3)</td>
<td>2 (6)</td>
</tr>
<tr>
<td>Other/no response</td>
<td>19 (28)</td>
<td>11 (32)</td>
</tr>
</tbody>
</table>

Table 4. Association between patient demographics and user groups.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Interested and engaged (n=68)</th>
<th>Interested but not engaged (n=34)</th>
<th>Refused (n=34)</th>
<th>Did not respond (n=63)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>Ref</td>
<td>.52</td>
<td>1.01 (0.97-1.05)</td>
<td>.55 (1.02-1.05)</td>
</tr>
<tr>
<td>Male</td>
<td>Ref</td>
<td>.52</td>
<td>1.01 (0.97-1.05)</td>
<td>.55 (1.02-1.05)</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>Ref</td>
<td>.52</td>
<td>1.01 (0.97-1.05)</td>
<td>.55 (1.02-1.05)</td>
</tr>
<tr>
<td>Other</td>
<td>Ref</td>
<td>.52</td>
<td>1.01 (0.97-1.05)</td>
<td>.55 (1.02-1.05)</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>Ref</td>
<td>.52</td>
<td>1.01 (0.97-1.05)</td>
<td>.55 (1.02-1.05)</td>
</tr>
<tr>
<td>BMI (kg/m$^2$)</td>
<td>Ref</td>
<td>.52</td>
<td>1.01 (0.97-1.05)</td>
<td>.55 (1.02-1.05)</td>
</tr>
</tbody>
</table>

$^a$OR: odds ratio.

$^b$All $P$ values were based on multivariable adjusted multinomial regression models, with the reference group being the interested and engaged group.

$^c$Not applicable.

Psychological Data

Overall, 4 standard screening psychological surveys were completed by 41 patients who were interested in using the digital tool. As a control, we used data collected from patients within our obesity service (n=633) who had completed these screening surveys previously.

Three-quarters of patients in the app group scored $\geq$10 on the PHQ-9 measure (28/37, 76%) compared to the control group (350/633, 55.5%). Just over half of patients (21/38, 55%) and 46.6% (294/633) scored $\geq$10 on the GAD-7 questionnaire in the app group and control group respectively. Eight of 41 (19%) patients in the app group and 166 of 633 (26.9%) in the control group endorsed thoughts about suicide or self-harm on the PHQ-9.

There were no statistically significant differences in scores between the app and control groups of patients (n=633) for Warwick-Edinburgh Mental Wellbeing Scale, PHQ-9, GAD-7, Dutch Eating Behavior Questionnaire–restrained eating, and Dutch Eating Behavior Questionnaire–external eating. However, compared with the controls, patients interested in using the app had significantly higher scores for the Dutch Eating Behavior Questionnaire–emotional eating ($P=.01$; Table 5).
Table 5. Scores for psychological screening surveys.

<table>
<thead>
<tr>
<th></th>
<th>WEMWBSa, median score (IQR)</th>
<th>GAD-7b, median score (IQR)</th>
<th>PHQ-9c, median score (IQR)</th>
<th>DEBQ-Rd, median score (IQR)</th>
<th>DEBQ-e, median score (IQR)</th>
<th>DEBQ-extf, median score (IQR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interested in the app (n=41)</td>
<td>39.5 (31.5-44.3)</td>
<td>10.5 (6-16.3)</td>
<td>14 (9.3-18)</td>
<td>29 (24-34.5)</td>
<td>44 (36-51)</td>
<td>31 (25.3-35.5)</td>
</tr>
<tr>
<td>Control group (n=633)</td>
<td>40 (33-48)</td>
<td>9 (5-15)</td>
<td>11 (6-17)</td>
<td>29 (23-34)</td>
<td>38 (26-50)</td>
<td>29 (24-35)</td>
</tr>
</tbody>
</table>

aWEMWBS: Warwick-Edinburgh Mental Wellbeing Scale.
bGAD-7: generalized anxiety disorder-7.
cPHQ-9: Patient Health Questionnaire-9.
dDEBQ-R: Dutch Eating Behavior Questionnaire–restrained eating.
eDEBQ-e: Dutch Eating Behavior Questionnaire–emotional eating.
fDEBQ-ext: Dutch Eating Behavior Questionnaire–external eating.

Metabolic Parameters
Sixty-two of 102 patients (60.8%) who were interested in the app and 29 of the 34 patients (85%) who refused the app had a screening blood test done by August 2021. The mean values of glycated hemoglobin, triglycerides, and low-density lipoprotein cholesterol are summarized in Table 6. There were no statistically significant differences between the metabolic parameters of those patients who were interested in and those who refused the offer of the app.

Table 6. Baseline blood test.

<table>
<thead>
<tr>
<th></th>
<th>HbA1ca (mmol/mol), median (IQR)</th>
<th>TGb (mmol/l), median (IQR)</th>
<th>LDLc-cholesterol (mmol/l), mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interested</td>
<td>39 (36-45)</td>
<td>1.8 (1.3-2.2)</td>
<td>2.6 (0.7)</td>
</tr>
<tr>
<td>Refused</td>
<td>37 (35-42)</td>
<td>1.8 (1.4-2.3)</td>
<td>2.9 (1)</td>
</tr>
</tbody>
</table>

aHbA1c: glycated hemoglobin.
bTG: triglycerides.
cLDL: low-density lipoprotein.

Engagement With the Gro Health App
Of the 68 patients who registered with the Gro Health app, 62 (91%) remained engaged at follow-up (defined as having opened the app or imputed data within the last month; data assessed in August 2021). Overall, engagement with the app was 60.8% among those who expressed an initial interest (62/102) and 31.2% (62/199) of patients who were offered the app. Overall, the mean duration of engagement with the app was 184.5 (SD 24.55) days. All patients selected a health goal, with a majority (67/68, 98%) selecting weight loss. All patients who engaged with the app also selected a health focus, which flagged the area of the app the user was currently engaged with, from the 4 therapy areas provided in the app. These included mental well-being (32/68, 47%) and nutrition (36/68, 53%).

Patients’ Input
To understand the thought processes of our newly referred patients regarding the use of digital apps as part of their clinical care, patient and public engagement workshops were held in January, April, and May 2021. Three main topics were discussed: (1) exposure to digital apps as part of weight management, (2) helpful features of existing digital apps, and (3) any desired features that would be helpful to future patients. The notes from the workshop are summarized in Textbox 1. The results from patient engagement workshops were used to create a bespoke version of Gro Health in order to provide a more tailored digital tool for this group of patients. Additionally, patients who were interested in using the Gro Health app were asked to complete a feedback form regarding their experience. Anonymized feedback was received from 11 participants. The reported reasons for discontinuation of the app included difficulty in quantifying the weight of food, problems integrating the app with other accessories, and forgetting to use the app. The most common goal that participants set on the app was weight loss. The features of the app that were most enjoyed were weekly educational lessons, downloadable behavior change activities and resources, and health tracking. Patients felt that food lessons improved what and how they eat. Most patients who responded (8/11, 73%) thought that the app was of high or very high credibility.

**Exposure to apps**
- The apps that were free were more desirable.
- The fact that the apps are not homegrown and not recommended by their physician and general practitioner makes choosing an app difficult. Recommendations of digital apps from health care professionals would help to improve confidence.
- Other departments such as physiotherapy have incorporated web-based materials (website and an app) in their services that boost patients’ confidence in their use.

**Helpful features of existing apps**
- Sets goals for them including carb counting, fat, and proteins
- Food tracking and respective nutritional information
- Glucose monitoring
- Explanation of food groups
- Conflates data from other apps

**What the patient would want in an app**
- They would want 1 app or product tailored to all needs
- Recommended menus
- Sends orders to supermarkets
- Simple
- Improved clarity in the instructions to patients
- Recommended or prescribed by their physician or general practitioner

**Discussion**

**Principal Findings**

We report on the first assessment of interest and engagement of patients awaiting input from a hospital-based obesity service with the digital tool Gro Health. There was significant interest from patients who were referred into NHS weight management services to use a digital tool to support their weight management journey. Emotional eating and higher BMI were associated with interest and engagement with the Gro Health app. This could be explained by an increased desire for additional support tools among those with higher likelihood of emotional eating and higher BMI. Men were less likely to engage with the Gro Health app than women. However, we did not identify any other predictors of patient interest in the digital app, such as ethnicity, age, or metabolic measures of glycemia and lipid profiles.

We identified from patient engagement workshops that given the plethora of health-based apps currently available, a recommendation for the use of specific digital tools, such as Gro Health, should ideally be provided by a health care professional, with clear instructions on its optimal usage. In addition, patients provided invaluable insight into features they would like to see in any digital weight management tool. This is extremely important, given the recommendation of Topol Review that patients need to be included as partners (and encourage cocreation) when it comes to health technologies [19].

It is important to highlight that 63 people (32%) did not respond to our invitation, and it is not possible to conclude whether they were not interested, did not receive the right information, lost the letter, or forgot to reply. As a learning point from this, a landing page (web page) for the digital tool Gro Health was created. This provides all the necessary information about the digital tool, and it registers the interest of potential users.

For those participants who engaged with the Gro Health app, there was a high engagement and retention rate, similar to other reported studies using the intervention [13]. To improve future user engagement with digital health care apps, it is important that we learn from the existing literature within the field. In a meta-analysis by Szinay et al [20], the factors associated with higher uptake of a health-related app were availability at low cost, awareness of the app, and recommendations by clinicians. Factors associated with higher user engagement included user guidance, personalization, statistical data on progress, and self-monitoring features [20]. In a recent meta-analysis by Spaulding et al [21], although increased health app engagement was associated with improved weight and BMI, the authors suggested that further research is required to further understand mobile health user engagement in both inpatient and outpatient setting [21].

In recent years, there has been a substantial acceleration in the uptake and engagement with health-related apps, generally, reflective of the increasing digitalization of the health care delivery. This recent health care digitalization revolution has been catalyzed somewhat by the COVID-19 pandemic that has necessitated fundamental changes in the delivery of health care, including widespread implementation of remote appointments between patients and their health care teams. Our current health care digitalization revolution within the NHS offers huge
potential for improvements in patient care and the efficiency of delivery of health care innovations. NHS obesity management is no exception. However, understanding the factors that predict disengagement with digital tools is important to optimize their future use and clinical utility within NHS-based clinical settings. Education of health care staff about the availability and benefits of digital health care tools is required to improve their uptake among patients, with clear instructions on their use and recommendations from a health care professional. A recent systematic literature review identified several sociotechnical factors that influence patients’ adoption of mobile health tools [22]. Some of the key findings from this comprehensive review were also seen as themes emerging from our patient engagement workshops, such as cost of the digital tools, incorporation into clinical pathways, and provision of appropriate health education and self-management.

This review provided a clear recommendation on a patient-centered approach that promotes patient adoption, with some of the key features such as fitting into patient’s overall treatment journey, inclusive design (especially for those users with less digital experience), comprehensive patient education and support, encouragement of the entire clinical team to use these tools, strong data ethics, and appropriate incorporation into health care policy [22].

Finally, we demonstrate the feasibility of the implementation of digital tool, Gro Health, to patients awaiting their first clinical appointment within our hospital-based obesity management team. Digital tools in weight management should not replace proper assessment and input from relevant health care professionals but rather augment traditional clinical care to optimize clinical efficiency in a novel, hybrid (blended) model of health care.

To fully embrace the digital health care revolution, its benefits, and huge potential, it is important that patients, their health care teams, and providers are involved in the creation of novel and bespoke digital health care tools for the future. As demonstrated from patient feedback, health care professional endorsement and patient cocreation are factors impacting any digital tool uptake. This highlights the importance of training and guidance for health care professionals to support patients with digital tools.

In addition, with so many digital tools available, it is key that tools demonstrate patient, clinician, and system benefits before adoption within health care systems. To this point, at the time of writing, Gro Health is the highest-rated digital health app as assessed by Orcha—reviewers of digital health apps on behalf of the NHS [23].

Digital tools such as Gro Health provide a foundation to support any unmet needs with education, behavioral support, and optimal user engagement that ultimately improves both the efficiency of health care delivery and, of course, patient outcomes. Future studies will assess the impact of such an initiative on patient-based outcomes and how these compare to traditional models in which there is usually very little or no patient contact and support from hospital-based clinical teams prior to their first clinical appointment.

Limitations
This was an observational study with a relatively small number of participants. A formal power calculation was not performed. Additionally, it was not possible to completely assess the reasons for lack of interest among the third of eligible patients who did not respond to any contact.

Conclusions
Emotional eating and higher BMI were associated with interest in the digital tool, Gro Health. Male gender was associated with reduced engagement with the app. There was no association between age or ethnicity and interest in the use of Gro Health app. Recommendations for the use of specific digital tools should ideally be provided by a health care professional, with clear instructions on its optimal usage. Additionally, patients should be involved in the cocreation of digital health tools.

Further research should evaluate the clinical impact of digital tools, such as Gro Health, as well as explore barriers and facilitators to engagement with digital tools in specialist weight management service.

Acknowledgments
This work was done as part of a Topol Digital Fellowship awarded by the Health Education England.

Conflicts of Interest
CS and AP are founders of Diabetes Digital Media Health, and MdlF and AK are employed by Diabetes Digital Media Health.

Multimedia Appendix 1
Screenshots of the Gro Health app and key features.
[PDF File (Adobe PDF File), 1238 KB - humanfactors_v10i1e41256_app1.pdf]

References


22. ORCHA. ORCHA. URL: https://appfinder.orchaph.org/ [accessed 2023-04-13]

Abbreviations

COM-B: capability, opportunity, motivation, behavior
GAD-7: generalized anxiety disorder
Implemntation of a Digital Health Tool for Patients Awaiting Input From a Specialist Weight Management Team: Observational Study

JMIR Hum Factors 2023;10:e41256
URL: https://humanfactors.jmir.org/2023/1/e41256
doi:10.2196/41256
PMID:37256653

©Petra Hanson, Charlotte Summers, Arjun Panesar, Alexandros Leonidas Liarakos, Dominic Oduro-Donkor, Danniella Whyte Oshodi, Luke Hailston, Harpal Randeva, Vinod Menon, Michaela de la Fosse, Amit Kaura, Emma Shuttlewood, Mark Loveder, Donna Poole, Thomas M Barber. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 31.05.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Original Paper

A Transgender Health Information Resource: Participatory Design Study

Brad Morse¹,², MA, PhD; Andrey Soares¹, PhD; Bethany M Kwan²,³, MSPH, PhD; Marvyn Allen⁴, MPA; Rita S Lee¹, MD; Kristen Desanto⁵, MSLS; Brooke Dorsey Holliman²,⁶, PhD; Kate Ytell⁷, MPH; Lisa M Schilling¹,², MSPH, MD

¹Division of General Internal Medicine, Department of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, United States
²Adult and Child Center for Outcomes Research and Delivery Science, University of Colorado Anschutz Medical Campus, Aurora, CO, United States
³Department of Emergency Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, United States
⁴One Colorado, Denver, CO, United States
⁵Strauss Health Sciences Library, University of Colorado Anschutz Medical Campus, Aurora, CO, United States
⁶Department of Family Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, United States
⁷Elevance Health, Denver, CO, United States

Corresponding Author:
Lisa M Schilling, MSPH, MD
Division of General Internal Medicine, Department of Medicine
University of Colorado Anschutz Medical Campus
8th Floor, Academic Office 1-Office 8219
12631 E 17th Ave
Aurora, CO, 80045
United States
Phone: 1 303 724 2254
Email: lisa.schilling@cuanschutz.edu

Abstract

Background: Despite the abundance of health information on the internet for people who identify as transgender and gender diverse (TGD), much of the content used is found on social media channels, requiring individuals to vet the information for relevance and quality.

Objective: We developed a prototype transgender health information resource (TGHIR) delivered via a mobile app to provide credible health and wellness information for people who are TGD.

Methods: We partnered with the TGD community and used a participatory design approach that included focus groups and co-design sessions to identify users’ needs and priorities. We used the Agile software development methodology to build the prototype. A medical librarian and physicians with expertise in transgender health curated a set of 97 information resources that constituted the foundational content of the prototype. To evaluate the prototype TGHIR app, we assessed the app with test users, using a single item from the System Usability Scale to assess feature usability, cognitive walk-throughs, and the user version of the Mobile Application Rating Scale to evaluate the app’s objective and subjective quality.

Results: A total of 13 people who identified as TGD or TGD allies rated their satisfaction with 9 of 10 (90%) app features as good to excellent, and 1 (10%) of the features—the ability to filter to narrow TGHIR resources—was rated as okay. The overall quality score on the user version of the Mobile Application Rating Scale was 4.25 out of 5 after 4 weeks of use, indicating a good-quality mobile app. The information subscore received the highest rating, at 4.75 out of 5.

Conclusions: Community partnership and participatory design were effective in the development of the TGHIR app, resulting in an information resource app with satisfactory features and overall high-quality ratings. Test users felt that the TGHIR app would be helpful for people who are TGD and their care partners.

(JMIR Hum Factors 2023;10:e42382) doi:10.2196/42382

KEYWORDS
lesbian, gay, bisexual, transgender, and queer; LGBTQ; transgender; mobile app; health information; participatory design; agile development; mobile phone
Introduction

Transgender Health Disparities and Inequities

The term transgender and gender diverse (TGD) refers to individuals whose gender identity does not align with their sex assigned at birth and includes individuals who identify as transgender, gender nonbinary, gender diverse, or gender-fluid [1]. The term transgender can be viewed in contrast to the term cisgender, which is used to describe those whose gender identity aligns with their sex assigned at birth.

The social determinants of health and structural stigma create health barriers for the TGD population. For example, 29% of the respondents of the 2015 US Transgender Survey lived in poverty compared with 12% of the US population [2]. In addition, the unemployment rate of respondents was 15% compared with 5% in the general US population [2]. As a result, the TGD community experiences many barriers to achieving health and well-being. Relative to cisgender individuals, transgender individuals experience disparities and inequities in all aspects of health (ie, mental, physical, emotional, and social), including poor overall health status, access to quality health care, and mental health, as well as an increased risk of substance use disorder, myocardial infarction, and sexual and reproductive health concerns [3-7]. Transgender individuals commonly experience trauma, abuse, and violence throughout their life span, leading to chronic stress [8,9]. Persons who are TGD experience many social determinants linked to poor health, including a lack of stable income and quality housing [10-12].

Stigma [13,14], ignorance [15], and discrimination contribute to poor health care access and poor care quality. Lack of social support, ranging from social exclusion and marginalization to poor social and family relationships, including rejection and family violence, contributes to stress and mental illness, including suicide [16]. Persons who are TGD are often poorly treated in the health care system and have difficulty finding TGD-competent and knowledgeable providers [15]. As a result, persons who are TGD may avoid routine care and are less likely to receive preventive services [17], less adherent to life-saving therapy [18-20], and more likely to experience denial of services from primary care to end-of-life care [21,22] and to be uninsured or have public insurance than persons who are cisgender [23]. In 2015, the American College of Physicians advocated for the creation of policies that would advance health equity among the lesbian, gay, bisexual, transgender, queer or questioning (LGBTQ) community—including the need for ongoing research on best practices for equitable health care [24]. Complicating the need for TGD-competent care is the fact that persons who are TGD experience unique needs related to gender-affirming health care as well as general health care and that their needs evolve across their life span, from youth and adolescence through midlife and late life [13,25-27].

Health Information Needs of Persons Who Are TGD

The existing literature shows that persons who are TGD and their caregivers most often seek information to (1) explore gender identity and coming out, (2) fill health and medical knowledge gaps, (3) seek support networks, (4) find TGD-competent providers, (5) find legal advice, and (6) find advocacy/political advice [28,29]. Legal issues important to health and well-being include protections, such as health insurance discrimination, and policies and procedures for changing gender information on legal forms of identification such as a driver’s license [30]. Persons who are TGD also seek information to build skills for communication with health care providers [25,31] and information regarding strategies and counseling services to build resilience, improve body image, and contend with other stressors [32]. In addition, the TGD community needs specific medical and health information regarding gender-affirming care, including surgery, mental health support, and hormone treatments and their impact [28,29,33-37].

Barriers to finding relevant health information include (1) a general lack of TGD health information [35], (2) not knowing the terms to use when searching [35], (3) often finding hateful content and misinformation [34], and (4) identifying credible and reliable sources of health and medical information. The TGD community relies heavily on the internet for all types of information [29]. Having access to credible sources of health information could serve to balance the increasing number of articles or posts that contain misinformation or are outdated [38,39]. Evans et al [28] highlighted the need for credible and trustworthy web-based content. Relevant TGD health information should be easy to access and broadly meet the diverse needs of the TGD community [34].

Transgender Health Information Sources and Accessibility

A 2012 review identified and categorized several websites created for the TGD community [40]. Within the health domain, HIV was the focus of most of the websites (n=17), followed by gender-affirming surgery (n=8), mental health (n=2), primary care (n=1), and sexual health (n=1). Over the last decade, with support from the National Library of Medicine, public and health sciences libraries have focused on cataloging information resources for the TGD community, including establishing transgender resource library collections and reference services [25,26,29,41,42]. The Lesbian, Gay, Bisexual, and Transgender Health Resources Guide of the University of Colorado Strauss Health Sciences Library [43] is one such collection.

Owing to the phenomenon of mobile apps for dating and social networking, their use among the LGBTQ community has been studied extensively to assess users’ risk of HIV and sexually transmitted infections [44-47]. A recent study by Akinola et al [36] assessed the barriers and facilitators for Black transgender women to use the mobile app technology for HIV self-testing and remote research participation. Reported facilitators included being more engaged and having increased self-agency, whereas barriers included inconsistent access to the internet and smartphones. Radix et al [48] concluded that the use of health ITs (HITs) provides opportunities to improve the quality of care for TGD individuals. Not only can HIT solutions be designed to offer education and support addressing the social determinants of health, but the community also favors these solutions [48].
Objectives
Our objective was to use participatory design methods to design a health information resource to support persons who are TGD in finding and using credible health information prioritized according to their needs. We selected a delivery method known to combat the disparities of the digital divide and known to be used by the LGBTQ community already—mobile smartphones [44-47,49,50].

Methods
Overview
This section describes the design and development of our transgender health information resource (TGHIR [tigger]) [51,52]. The TGHIR platform consists of three main components (Figure 1): (1) curated health and wellness information content (henceforth, TGHIR resources), such as websites, documents, videos, and other consumer health apps; (2) a back-end system, including a database to store user data and TGHIR resources, a search engine, and a communication platform that allows users to send messages to the development team (eg, comments/feedback, suggestions of new content, and reports of offensive or inaccurate content); and (3) a front-end mobile app (henceforth, TGHIR app) that allows users to create accounts, search for information, and access/view the curated TGHIR resources. We did not create any of the TGHIR resources and, instead, incorporated links to freely available resources that can be accessed via the internet. The TGHIR resources were used as input to populate the database and as seeds to create the search engine.

Figure 1. Overview of the transgender health information resource (TGHIR) components.

Study Design
We based the design and development of the TGHIR app on a conceptual framework (Figure 2) that recognized the importance of participatory design, which is a form of user-centered design that focuses on designing with end users and not merely for end users [53]. Participatory design prioritizes users and allows for their direct participation in the design process through decision-making, going beyond the role of users as consultants. Direct involvement in the entire design process often leads to greater satisfaction with both the process and the outcome [51,52]. Participatory design also represents a key strategy for designing for dissemination, sustainability, and equity, attending to potential factors that may influence widespread adoption and equitable access to the TGHIR app [54].

We applied the basic stages of participatory design by Spinuzzi [55]—stage 1: initial exploration of work; stage 2: discovery processes; and stage 3: prototyping—and implemented a participatory design process described by Schnall et al [56]. We applied the 4-phase approach by Schnall et al [56] to the design and evaluation of the TGHIR app (Figure 3). In this paper, we sequentially summarize the methods and results by phase as the results from subsequent phases inform later phases.
Ethics Approval

The project was approved as exempt human participant research by the Colorado Multiple Institutional Review Board (protocol 19-1562).

Collaborators

Overview

Collaborators involved in the development, design, and evaluation of the TGHIR included our research team; an LGBTQ community-based advocacy organization (One Colorado); our
Community Advisory Board (CAB); and research participants, including focus group (FG) participants, design session co-designers, HIT experts, and test users. One Colorado is a leading advocacy organization in Colorado dedicated to advancing equality for LGBTQ Coloradans and their families [57]. The research team consisted of individuals with clinical, informatics, library/information science, dissemination and implementation science, and health service research expertise.

In partnership with One Colorado, we established and engaged a CAB. The CAB included 20 individuals. They were 60% (12/20) community members and 40% (8/20) research team members. CAB community members included 7 people who identified as TGD; 4 parents of TGD adolescents; and 3 clinicians with expertise in TGD health care, including primary care, surgery, and mental health. Some CAB members belonged to more than one category. We interviewed everyone to confirm their interest in and commitment to the project. The CAB was engaged throughout all phases of the design and development process.

**Research Participant Recruitment**

In collaboration with One Colorado, research participants were recruited using a series of Facebook recruitment posts in private transgender groups and physical fliers posted at the Integrated Transgender Clinic [58] at the University of Colorado Anschutz Medical Campus. People expressing interest in the project first met via Zoom (Zoom Video Communications) [59], a communication and collaboration platform, with the project manager, who also co-led the design sessions, to confirm interest in and fit for the project. Our goal was to recruit those who were TGD or TGD allies and who wanted to contribute to a participatory design process to design a TGD health information resource.

Overall, 42 TGD individuals responded to the recruitment advertisement, and 32 (76%) participated in phases 1 to 3. A total of 10 individuals were not eligible: 2 (20%) because of gender identity and age and 8 (80%) because of scheduling conflicts. A total of 81% (26/32) participated in an FG and 59% (19/32) participated in one or two design sessions, including 13 that attended both an FG and a design session. All 32 of the participants self-identified as either transgender, non-binary, gender queer, or other TGD category. The race and ethnicity reported by participants was mainly White (21/32, 66%) and other categories (11/32, 34%) included Hispanic, African American, Asian, Native American, and multiple races. Most participants were between 18 to 40 years of age, 72%, (23/32). In total, 2 HIT experts in usability were recruited from our medical center to participate in phase 3. A total of 13 test users not involved in phases 1 to 3 were recruited for phase 4 in the aforementioned manner. In total, 92% (12/13) of the test users identified as either transgender, nonbinary, or genderqueer. Due to low sample size we cannot report race, ethnicity or age ranges.

**Phase 1: Health Information Needs Assessment**

Phase 1 was dedicated to establishing a relationship between the CAB and the research team, eliciting CAB member insights into the proposed methods, and identifying TGD-related health information needs and sources.

**CAB Engagement**

**Overview**

The project started with a 4-hour kickoff meeting with the CAB and the research team using Liberating Structures [60], which are meeting strategies and structures that replace traditional top-down meeting practices with whole-group interactions. We used the Liberating Structures Purpose-to-Practice exercise to generate shared purpose, principles, participants, structure, and practices, which helped define the research team’s and the CAB’s responsibilities and approach to the work. The CAB also reviewed a draft of the FG guide and recruitment fliers at the CAB kickoff meeting.

**CAB Kickoff Insights**

The CAB recommended edits to the FG materials and overall approach, including recommendations on (1) term use (eg, transgender and gender diverse rather than transgender and nonbinary) and (2) specific health topics (eg, mental health support and finding clinicians) that we should inquire about in the FGs. The CAB also ensured that we understood the history of challenges and dissatisfaction that the TGD community has had with the health care system in general. This included the challenges that the TGD community faces in finding clinically (ie, knowledge of evidence-based health care for TGD people) and culturally (ie, trans-friendly) competent care. The CAB stressed that a common problem in the TGD community is that health care professionals frequently attribute all medical conditions to being TGD and that TGD individuals must manage the same general health and medical needs as the cisgender community (eg, cancer prevention and broken bones). These insights not only improved our FG materials but also allowed us to be sensitive to the frustrations the community has with health care providers and systems, which to them the research team represented.

**FG Engagement**

**Overview**

We conducted 4 FGs with 26 participants (n=8, 31% in FG 1; n=6, 23% in FG 2; n=6, 23% in FG 3; and n=6, 23% in FG 4) following a phenomenological approach to design, conduct, and qualitatively analyze a TGD health information needs assessment [61]. We developed a semistructured FG guide using essential questions [61] to understand health information–seeking behavior (ie, how and where) and the types of health information sought (ie, what). We conducted web-based FGs using Zoom. FGs were recorded and professionally transcribed for analysis. We performed rapid analysis [62] of the transcripts to allow us to quickly use the information in the next resource information app development phase.

**Results: FG**

The FGs ultimately provided limited information about the specific types of health information needed by the TGD community as the FG participants organically oriented their discussions to the need for clinically and culturally competent care. However, 2 key insights did emerge from FG analyses. First, participants reported that their health care providers often
did not have answers to TGD-specific health questions and advised them to seek information themselves using the internet. Second, many participants indicated that they relied heavily on peer-to-peer social media (eg, Reddit and private Facebook groups) for health information. However, this information was often felt to be difficult to interpret in the context of their specific gender identity and required vetting on their own. Taken together, these insights suggested that there was an unmet need regarding finding credible, personally relevant health information, and thus, an app such as the TGHIR app could be useful.

**Health Information Source Identification and Curation**

*Overview*

Additional strategies for identifying the types of health information often sought by TGD individuals and the corresponding sources of that information included performing literature and internet searches and seeking input from the CAB. CAB members emphasized the need for information about seeking health insurance, hormone therapy, and legal resources (eg, updating driver’s licenses and permission to be at the bedside of a hospitalized partner) and locating competent clinicians. The CAB also emphasized that the information resources should be sensitive to terms and stigmatizing language, accessibility (disability and non-English language), and inclusivity. In parallel, the research team’s medical librarian (KD) performed a literature search on a medical bibliographic database (ie, PubMed) for research articles on the health information-seeking behavior of TGD people. This process identified the categories of information most frequently searched on the web (eg, hormone therapy, health insurance, mental health support, and surgery options) [63-66]. Next, the librarian performed a Google search to identify health information resources freely available to the public in each of these categories. A set of search terms (Multimedia Appendix 1) was developed based on published terminology to identify LGBTQ-related information [67]. CAB members reviewed and provided insights on the list of health information resources, specifically focusing on issues of credibility and inclusivity. These recommendations were aligned with the research team, and the CAB provided critical suggestions on how to carry them out. It was deemed important that users of the TGHIR app understand how to assess web-based health information for credibility as it would not be sustainable to have a medical librarian or health professional review every potential resource.

We asked the CAB to review available layperson credibility assessment tools (Multimedia Appendix 2) that could be incorporated into the app and help users make judgments for themselves. The CAB agreed that the simpler of the tools, *Trust it or Trash it* [68], would work best for determining credibility. The CAB asserted that this tool was informal, the format was easy to follow, and it was more accessible than a MEDLINE tutorial on how to evaluate health information found on the internet [69]. The CAB also wanted to ensure that the app was inclusive regarding the broad range of gender-diverse identities. For them, this meant using terms acceptable to the TGD community and including a broad range of information for TGD persons across their life span. CAB members suggested search tags and category labels (described in the following section) to ensure the use of culturally acceptable and commonly used terms to find resources within the app. The librarian and clinician research team members reviewed the initial sources for credibility before finalizing the resource list, tags, and category labels.

**Results: Health Information Curation**

A total of 97 credible health information resources in 16 topical categories (Multimedia Appendix 3) were identified and cataloged for this project. The list of resources and categories was often revised and updated as new resources were suggested or discovered throughout the project. The medical librarian ensured that all the web links were active and provided access to the expected content. Each information resource was cataloged regarding name/title, authoring organization, topic (eg, health, mental health and social support, or legal and financial), potential search tags, and category (eg, surgery and health care rights) for storage in the TGHIR database.

**Phase 2: Prototype Design, Evaluation, and Usability Test**

*Overview*

In phase 2, we identified and prioritized features and designed prototypes (mock-up screens) using Justinmind [70]. In a series of web-based sessions, participatory co-designers considered how they would want to access and use TGD health information resources via the TGHIR app. *Table 1* provides an overview of the 4 web-based design sessions that occurred in phases 2 (sessions 1-3) and 3 (session 4), including the number of participants and the primary methods used.

![Table 1. Overview of the design sessions conducted in phases 2 and 3.](https://humanfactors.jmir.org/2023/1/e42382)

**Session** | **Research participants, N** | **Goals/tasks** | **Design approach**
--- | --- | --- | ---
1: feature requirement exploration | 4 | Exploration of mobile app features liked and used and value proposition generation | Co-design session
2: feature prioritization | 13 research participants and 9 CAB members | Feature prioritization | Kano customer satisfaction survey
3: app design and esthetics | 7 | Wireframing and prototyping; determining the visual esthetic of the app | Participatory design and iteration; semistructured interview and idea generation
4: expert usability testing | 2 (expert evaluators) | Feature usability testing | Heuristic evaluation; cognitive walk-through and System Usability Scale scores
**Design Session 1: Feature Requirement Exploration**

**Overview**
The first design session was exploratory and represented stage 1 of the participatory design approach by Spinuzzi [55]. We asked participants to free list app features that they found satisfying and sort the features into thematic categories (e.g., privacy and user interface). Free listing is a fast way to generate many ideas in a short period [71]. During the exercise, participants created sticky notes with desired features and then moved the sticky notes between quadrants, conferring and discussing their opinions with the other participants. We also conducted a value proposition exercise to understand end users’ expectations of using the TGHIR app [72]. We asked participants to consider the perspectives of other potential end users using a templated value proposition statement (i.e., For people who ___, this mobile app is ___ that will provide ___).

**Results: Design Session 1**
A total of 4 participants generated 35 potential features and 14 value propositions. The research team extrapolated additional features from the value propositions. Figure 4 shows an example of a value proposition. The research team then grouped similar features and removed any duplicates to generate a final list of 23 features (Textbox 1), which was used as input for design session 2.

![Figure 4. Sample feature value proposition statement from design session 1.](image)
Textbox 1. Features classified using the Kano Model of Customer Satisfaction.

**Must-be (requirements that a user expects to be implemented and that the user would be dissatisfied if they were not available)**
- External links to resources in general and links to hormone therapy mentioned specifically (external resources)

**One-dimensional (requirements are related to the quality of a feature or service such that greater quality is correlated with greater satisfaction)**
- Search by topic and subcategory functionality (user interface)
- External links to mental health services (external resources)
- External links to community services (external resources)
- Contact us functionality (reviews)
- User interface simplicity (user interface)
- Account settings (user interface)
- Use of in-app pronouns tailored to the end user (pronouns)
- Remembering user settings (user interface/privacy)
- Culturally relevant language (user interface)
- Accessibility options (user interface)
- Removal of inappropriate posts (reviews)
- No data sharing with third parties (privacy)
- Protection of credentials for logging in (privacy)
- Data security (privacy)

**Attractive (requirements may not be expected or expressed by a user but would make them satisfied if they were implemented)**
- Search by typing (interface)
- Ability of users to suggest new information sources (user interface/reviews)

**Indifferent (a user has a neutral opinion on whether a feature is implemented)**
- History of viewed content (user interface)
- External link to pronoun tester (pronouns)
- Gender identity filter to support searching resources (user interface)
- Resource synopsis (external resources)
- Review of resource information credibility (user interface)
- News feed (user interface)

### Design Session 2: Feature Prioritization

#### Overview
In design session 2, stage 2 (discovery) of the participatory design model by Spinuzzi [55], we asked participants to prioritize the 23 app features identified in design session 1 [55] using the Kano Model of Customer Satisfaction exercise [73,74]. The Kano Model classifies features as must-be, one-dimensional, attractive, and indifferent (Textbox 1). The Kano survey was completed via REDCap (Research Electronic Data Capture; Vanderbilt University). To allow the research team to better understand the justification for feature prioritization, several participants took part in a web-based facilitated exercise to prioritize features after independently completing the Kano survey. The web-based session allowed participants to use a bullseye visualization to prioritize features and prompted discussion about their choices.

### Results: Design Session 2
A total of 22 participants, including 9 CAB members, completed the Kano survey. Textbox 1 lists the features in order of prioritization according to the 22 Kano survey responses.

The bullseye exercise resulted in the image shown in Figure 5. Owing to the interactive nature of this activity, participants only discussed and arranged 65% (15/23) of the features. We compared the prioritization from the bullseye exercise with the classification from the Kano survey. The features in the highest-priority center of the bullseyes were links to references, no data sharing with 3rd parties (eg, not selling personal information), and “contact us,” all of which were classified as must-be or one-dimensional in the Kano results. The multilingual option and compatibility with assistive technology features were both in the center of the bullseye and also ranked as One-Dimensional; unfortunately, these were out of scope for this project’s funding.
Design Session 3: App Design

Overview

Design session 3, stage 3 (prototyping) of the participatory design model by Spinuzzi [55], was the wireframing and prototyping of the TGHIR app features. In this session, participants iteratively envisioned and informed the various features of the TGHIR app using midfidelity mock-up screens (Figures 6A and 6B). The prototypes were created by the design session moderator using the Justinmind prototyping tool [70]. We held 2 sessions covering the account creation process, the menu page, and the main information resource screens. We also shared an early prototype with the CAB to collect insights and feedback.
Results: Design Session 3—App Design
The 9 participants felt that all TGHIR app design elements should support easy access to the TGHIR resources and quick assessment of credibility. Owing to concerns about privacy and potential misgendering raised by the design participants, the account settings and use of in-app pronouns were discarded from the final list of features. This decision resulted in no use of personal profiles with self-identified pronouns or identifiers such as names and birthdates. The only personal information used by the system was the email address used to create a user account and log in to the system. However, after logging in, only the autogenerated hash ID would be used to record user activities, such as indicating resource likes and bookmarking resources.

Participants also felt that it was important to visually communicate the credibility of the TGHIR resources by displaying the content creator’s logo. Participants desired that the TGHIR Resource Card display the title, a short description, and a leading image and provide a preview of the resource, allowing them to choose whether to access it. The final version of each type of card varied slightly from the midfidelity examples. For instance, Figure 6C shows the final and enhanced design of the resource cards. Finally, some participants proposed changing the TGHIR app icon to a non–TGD-associated image such as a calendar.

CAB feedback also contributed to the revision and redesign of the app features. For instance, an earlier prototype forced users to immediately label themselves as either transmasculine or transfeminine upon entering the app in response to earlier input where users only wanted to see information relevant to their gender identity. This was ultimately deemed more harmful than helpful because of the breadth of gender identity diversity within the TGD community.

Design Session 3: Esthetics
Overview
One participant, a student of graphic art and user design, engaged in a 1:1 design session to provide input on the TGHIR app esthetics. The design session was conducted as a semistructured interview, and the participant was asked to provide their opinions and suggestions regarding the app’s user interface and its potential impact on user experience. They contributed to five topics related to esthetic appeal: (1) color palette, (2) font type, (3) name of the TGHIR app, (4) TGHIR app logo, and (5) use of backgrounds and images.

Results: Design Session 3—Aesthetics
This design session generated important feedback and insights to inform the final appearance of the app layout, color, and font scheme. The participant suggested blue hues for background, cards, and buttons (eg, “blue is a color that feels relaxed”) but recommended that we not use the Flutter (Google) default blue for the header and foot bar as it was similar to Facebook’s blue hue. Thus, to avoid similarities or confusion with commonly used social apps, we selected yellow for the app’s header and footer. It was also recommended not to use any background images or animations because of potential loading performance issues and user frustration.

Another suggestion was “making timely resources pop out on the category page so people can see it quickly if they need it.” This suggestion led to the creation of the categories of interest feature, where end users could select their preferred categories of interest and the TGHIR app pinned the categories to the top
of the Health Resource Cards page using a different background color to indicate pinned categories. For instance, selecting the categories hormone therapy and local resources turned the blue category cards into orange cards (see the example in Figure 7A). This option would allow users to have easy access to preferred categories. The TGHIR app offered two ways for end users to select preferred categories: (1) by clicking on the shortcut icon in the footer (first icon on the bottom left; Figure 7A) to open the categories of interest page where checkboxes can be used to indicate a preference and (2) by clicking on the pushpin icon that is available when viewing the list of resources in a specific category (icon on the right of the category title; Figure 7B). An overview of the app is provided in Figures 7A, 7B, 7C, and 7D. The app layout has 3 main areas. The first is the header, which includes the title of the page being displayed; the overall menu (hamburger menu icon; A and D); icons for features such as returning to the previous screen function (back icon; B and C) and marking the current category as preferred (pushpin icon; B); and icons to like, bookmark, copy, and share resources (C). The second area is the body, which is used mainly to display the list of category cards (A), the list of resource cards (B), the content of the resource (C), and the list of popular resources (D). The third area is the footer, which displays the icons to access features such as category selection, search, filter, settings, and contact us.

**Figure 7.** Screenshots of the main transgender health information resource (TGHIR) app features.

![A: View list of categories](image)

![B: View list of TGHIR resources in a category](image)

![C: View a TGHIR resource](image)

![D: View most liked, bookmarked, and viewed resources](image)

**Phase 3: Mobile App 1.0 Build and Release**

**App Development**

In this phase, we reviewed the features and prototypes defined in the previous phase and translated them into requirements for developing the TGHIR app. We used Flutter [75], an open-source user interface development kit by Google, to develop a cross-platform mobile app for both Android and iOS devices. The database uses Firebase Firestore (Firebase Inc) [76], a NoSQL cloud database; the search engine uses Amazon OpenSearch Service (Amazon Web Services) [77]; and the email system uses SendGrid [78], a communication platform for email transactions.

We adopted the Agile software development methodology [79-81], an effective and efficient method to develop software that uses an iterative approach, incremental development, continuous value delivery, and user feedback [82]. On the basis of the features implemented in a sprint (2-week development cycle), the development team performed functional, interface, performance, system, service, and security testing of the TGHIR app on both Android and iOS devices. Examples of screens with the main features developed are shown in Figure 7.

**Search Engine Development**

With the goal of enhancing access to relevant TGD health information, we created a search engine that indexed information from the librarian, identified TGHIR resource pages, and linked subpages within the same domain. The URLs for the list of TGHIR resources were used as seed pages for the search engine. The search feature in the app allowed a user to type a search term and receive in return a list of pages that contained the searched term. Users could use wildcards to search for alternate spellings and variations on a root word.

**Design Session 4: Expert Usability Testing**

**Overview**

During this phase, we conducted design session 4 (expert usability testing), which provided insights and feedback to revise the app features with the goal of enhancing usability. This session consisted of experienced health information technologists, referred to as expert evaluators, with backgrounds...
in app design and integration. The TGHIR app version 1.0 was uploaded to the Google Play Store and the Apple Store Connect (TestFlight) so that the expert evaluators could install and test the app during the session. Each of the experts completed a cognitive walk-through [83] and was asked to complete 10 tasks associated with the TGHIR app features to identify any usability problems. Expert evaluators independently responded to item 3 on the System Usability Scale (SUS) [84,85], “I thought the feature was easy to use,” for each of the 10 app features/tasks (Table 2). We selected this single item to minimize participant response burden and optimize face validity.

Table 2. Feature and task evaluation by experts during phase 3 and by test users during phase 4.

<table>
<thead>
<tr>
<th>Feature/task evaluation</th>
<th>Phase 3: design session 4 (expert evaluators; n=2)—SUS item 3</th>
<th>Phase 4: usability and quality evaluation (test users; n=13)—SUS items 3 and 8</th>
<th>Category score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Create an account</td>
<td>Strongly disagree/disagree, n 0 Neutral, n 0 Strongly agree/agree, n 2</td>
<td>Estimated SUS numeric score 78.42 Category score Good</td>
<td></td>
</tr>
<tr>
<td>Select preferred categories</td>
<td>2 0 0</td>
<td>77.74 Good</td>
<td></td>
</tr>
<tr>
<td>Find a specific item using the category cards</td>
<td>0 0 2</td>
<td>76.95 Good</td>
<td></td>
</tr>
<tr>
<td>Use filter to narrow resources</td>
<td>1 0 1</td>
<td>65.37 Okay</td>
<td></td>
</tr>
<tr>
<td>Like a resource</td>
<td>0 0 2</td>
<td>80.01 Good</td>
<td></td>
</tr>
<tr>
<td>Bookmark a resource</td>
<td>0 0 2</td>
<td>76.09 Good</td>
<td></td>
</tr>
<tr>
<td>Search to find a specific resource</td>
<td>0 1 1</td>
<td>68.46 Good</td>
<td></td>
</tr>
<tr>
<td>Send a message to developers</td>
<td>0 0 2</td>
<td>82.34 Excellent</td>
<td></td>
</tr>
<tr>
<td>Share a new resource for the community with the developers</td>
<td>0 0 2</td>
<td>79.25 Good</td>
<td></td>
</tr>
<tr>
<td>Find the most liked resource</td>
<td>0 0 2</td>
<td>85.42 Excellent</td>
<td></td>
</tr>
</tbody>
</table>

*SUS: System Usability Scale.

Results: Design Session 4

The results from our 2 expert evaluators showed that most app features were easy to use (Table 2). The selection of preferred categories, feature 2, was not perceived as easy to use by either expert evaluator. Neither expert evaluator saw the option of selecting a category of interest as obvious or intuitive; the menu option was buried, and the pushpin option was overlooked. Owing to the comments on difficulty finding and accessing some features, the development team created the footer with shortcut icons to facilitate access to the important features, such as category selection, search, filter, settings, and contact us interface.

One expert evaluator was concerned that the filter (feature 4) was difficult to use and end users would find it difficult to understand how the filter feature returned results. Specifically, the filter was built to facilitate searching for resources across multiple domains and categories. For instance, a user can select a broad domain, such as legal and financial topics, and be presented with categories that have information related to the selected domain (ie, health insurance, health care financing, health care rights, and legal resources). The filter also suggests broad categories based on a user’s narrow input. For example, a user selecting the narrow topic health insurance would be reminded, based on filtering, that the app has insurance resources related to legal and financial topics. This approach can help users learn about different facets of the transition process.

Phase 4: Final Assessment—Usability and Quality Evaluation

In phase 4, test users installed and used the released TGHIR mobile app on their smartphones for a period of 4 weeks and participated in 3 evaluations. First, we conducted cognitive walk-through interviews within 2 days of the TGHIR app being installed on the test users’ phones to understand the performance of all the features. Test users were asked to try the same 10 features/tasks that the experts had evaluated (Table 2). To assess feature usability, test users were asked to respond to items 3 and 8 on the SUS [85]—I thought the feature was easy to use and I found the feature very cumbersome/awkward to use, respectively [86]—for each of the 10 features via a REDCap survey. The SUS is unidimensional and only measures 1 construct, that is, perceived usability. It has been shown that collecting responses to items 3 and 8 is 96% accurate in assessing system usability while also decreasing participant burden [86].

Test users completed the user version of the Mobile App Rating Scale (uMARS) via REDCap at 2 and 4 weeks after app installation (T1 and T2, respectively). The uMARS is a widely used scale for evaluating the quality of mobile health (mHealth) apps [87]. The uMARS is a modified version of the Mobile App Rating Scale (MARS) that is designed to be used by end users of a mobile app without training or expertise in mHealth technology or in the related health field. The uMARS contains 16 items assessing 4 dimensions of objective quality—engagement, functionality, esthetics, and information—and includes 4 items for the dimension of subjective quality. We also used 3 optional items for the...
dimension of perceived impact. All uMARS items are assessed on a 5-point scale (1=inadequate, 2=poor, 3=acceptable, 4=good, and 5=excellent) [87]. The MARS and uMARS have been shown to have excellent internal consistency, with high individual Cronbach α values for all subscales, and are valid, reliable, and accurate in measuring health app quality [87,88]. The MARS scores have also been shown to correlate with mHealth app revenue, monthly active users, and user downloads [89].

Results

A total of 13 test users participated in the final usability and quality assessment of the TGHIR app version 1.0. The SUS assessment showed good to excellent usability for all features except the Use filter to narrow resources (Table 2). The cognitive walk-through interview results showed that the TGHIR app had high usability overall. Test user TGU104 said the following:

I love this app. My main feedback that I see growth in is a tutorial that does an overview...this is groundbreaking for there to be one place to actually find these links.

Once this test user realized that the TGHIR resources were credible health resources, they were excited and spoke highly of the TGHIR app. Another test user, TGU105, said the following:

I really like the way the cards look, and obviously I like the way the search function works, but you may want to find a way to integrate them.

In general, most of the feedback received was positive, with suggestions on how to make tweaks for increased usability. Users indicated that the filter worked and was relatively easy to find within the app (owing to the recognizable filter icon), but there was some confusion regarding the labeled parts of the feature (broad domains and narrow topics). A test user suggested that we add filters that would work to separate TGHIR resources by gender identity. Note that, upon sharing an early prototype of the TGHIR app with the CAB, the CAB raised concerns about a feature that required users to first select a gender identity, which the CAB felt might inappropriately force users to put themselves into a predefined category. Therefore, this feature was ultimately not incorporated into the app.

The results of the uMARS assessment are presented in Table 3. The overall mean uMARS objective quality score at the first evaluation (T1) was 4.13 (SD 0.29), indicating good overall TGHIR app quality. At the second evaluation (T2), the overall mean uMARS objective quality score increased to 4.25 (SD 0.35), indicating good overall TGHIR app quality. It appears that with continued use of the TGHIR app, test users perceived the app to be of better quality. The subjective quality rating was 3.75 (SD 0.83), and the perceived impact rating was 4.45 (SD 0.40). The information rating at T2 was 4.75 (SD 0.16), which is near excellent.

Discussion

Principal Findings

Through the process of participatory design [55], we successfully cocreated a TGHIR delivered as a mobile app for and with people who are TGD to provide access to credible health information resources. The participatory design process was guided by our conceptual framework, the multiphase methodology by Spinuzzi [55], and principles of community engagement. A partnership with the CAB and insights from research participants confirmed the TGD community’s need for accessible and credible health information. It also illustrated their frustrations and barriers to obtaining credible health information. Barriers to obtaining credible information include a lack of knowledge by clinicians; the lack of information on the internet and the inability to find and identify it even when it does exist; and, finally, the lack of empirical evidence for many of their questions. For the TGHIR app to be of value, it had to achieve 2 goals. First, the app should direct people to credible, useful information, and second, the app should be easy to use and satisfying to the user. After 4 weeks of use, users gave the TGHIR app an overall objective quality rating of 4.25, a subjective quality rating of 3.75, and a perceived impact rating of 4.45 out of 5.00 on the uMARS scale, which equates to a good-quality mobile app. Most importantly, the information category had the highest uMARS score of 4.75 out of 5.00. The uMARS scores indicate our success with both goals; a usable...
app and valuable information. The overall quality rating improved during the 2 weeks of use, suggesting that users gained familiarity with the app features. Mobile apps that have overall uMARS ratings of good quality are more likely to be adopted by the intended users than poor-quality apps [87,89]. Our 4-week engagement score was 3.98, and although this did increase from week 2 to week 4, it is not clear how this score would have changed over a longer time period. A review of publicly available asthma apps found that app quality varied and that engagement was often the lowest-scoring dimension [90]. Our TGHIR app scores were better than the asthma app scores that had average objective and subjective quality scores of 3.17 (range 1.54-4.55) and 2.65 (range 1.00-4.50), respectively. Our scorers were also the end users themselves, which may be better correlated with end-user use.

The intensity of engagement is interesting to consider for mHealth apps that seek to change user behavior, such as apps to help patients manage chronic diseases, where long-term and frequent engagement may be critical to improve outcomes [91,92]. Long-term engagement with diabetes management apps has been limited [93], and experts have suggested that virtual coaching [94] along with wellness and chronic disease app use may enhance engagement and outcomes. A recent paper exploring the value of mHealth apps for patients and how they can be incorporated into traditional medical care delivery first classified apps into four categories: (1) aiding diagnosis or decision-making, (2) improving outcomes through better disease management, (3) stand-alone digital therapeutic devices, and (4) primarily delivering education. [95] The TGHIR app falls best into category 4, as it was designed to provide information to the TGD community, and in that sense, it may assist with decision-making (ie, Should I go on testosterone or not?) and condition self-management but only by providing information and improving one’s knowledge. Information is essential for education, but we did not create specific educational materials, although some resources available through our app may have been developed with educational intent, nor was the TGHIR app a diagnostic or decision support app for a specific concern, such as whether a skin lesion is cancerous or to provide triage advice.

A recent randomized controlled trial of an mHealth app designed to deliver information to educate patients regarding knee pain showed that it increased a patient’s disease-related knowledge [96]. Others have also suggested that informational/educational apps may lead to improved processes and outcomes for diseases and conditions such as heart failure, inflammatory bowel disease, recent tonsillectomy, and various cancers [97-99]. Others have proposed that informational/educational apps may lead to improved patient-provider communications and improved decision-making, which is something that we would like to assess in future research.

The TGHIR app benefited from a research team that included clinicians who were experts in evidence-based transgender health care and a medical librarian who could search, identify, and deem credible the information resources made available via the app. Manually maintaining credible health information is a resource-intensive endeavor. Automated methods to find information on both peer-reviewed medical manuscripts and non-peer-reviewed materials would be valuable, but this also needs to be paired by credibility assessments and methods to tag content so that the TGHIR content can be searched for and found by users. Community crowdsourcing to create useful and credible health information resources has been found to have more reliability when professionals or experts are present in the process of content creation. Many concerns exist regarding the expertise (ie, no medical or health education expertise) and intent of the authors (ie, want to sell a product) [100,101]. Whether community users can judge credibility is also debated. There is research suggesting that people often report judging source credibility, but observational studies suggest otherwise [102-104]. Owing to concerns expressed with finding credible information and concerns about author intentions, we did create a contact us feature allowing users to recommend additional resources and removal of content deemed noncredible or offensive. We built a dedicated PubMed or MEDLINE search for use with the MedlinePlus application programming interface [105] and a search feature for ClinicalTrials.gov, but the prebuilt searches were not incorporated into the final TGHIR app because of time constraints. A link to the “Trust It or Trash It” resource [68] was also not incorporated into the prototype app. Both are planned for future versions pending additional funding.

There is a clear need and value proposition for the TGHIR app. The information provided by the app is not intended to replace competent clinical TGD health care and is intended to assist laypersons so they have the information needed to be engaged and informed participants in their health and well-being. Some participatory methodological insights were generated during the project and may be helpful to other researchers. Our perception was that the TGD participants were grateful for the opportunity to take part and be involved in something that was being designed for them and with them. Several participants went beyond our requested involvement to offer advice on graphic design and other areas in which they had expertise.

**Limitations and Strengths**

Limitations included a focus on users based in Colorado and challenges in recruiting and, thus, co-designing with racially and ethnically diverse individuals. Our medical librarian did identify health information resources in Spanish, and our back-end health information resource database is designed to store the language of the resource, but we had limited resources for providing the TGHIR app interface in languages other than English and providing access to non–English-language resources. Despite striving to recruit demographically diverse co-design and end-user testing participants, we had difficulty recruiting TGD people of color and from rural settings. Although 34% (11/32) of our participants indicated that they were Black or African American, Asian, Native American, and of multiple races, we were not able to recruit any Black or African American test users. We can only hypothesize why this was difficult, including reasons such as Colorado’s demographics and the multitude of political issues necessitating this community to be active advocates at the time of this project. We expect that the TGHIR app would be useful to people of color and from rural settings regarding the user interface and functions, but it remains a question as to whether the information would have been rated...
as highly by these underrepresented groups. Some of the information resources were local TGD resources, which were more likely to be in the Denver metropolitan area and, therefore, might be deemed less valuable to those residing outside the metropolitan area. It is also true that rural areas are less likely to have the same extent of LGBTQ+ resources as urban areas, which is not a fault of the TGHIR app. The age distribution of our participants, 18 to ≥50 years, is a strength of this study.

Conclusions
Using methods of participatory design with the TGD community and in partnership with a CAB, we were able to co-design and develop a health information resource delivered via a mobile app for persons who are TGD and their care partners. Users felt that this app would be beneficial to them and that it provided needed information. A health information app is only as good as the information it makes accessible, and ongoing updating and maintenance of information resources in any information app is a challenge. Next steps include work to automate/semiautomate methods to identify relevant and credible information and testing for clinical effectiveness, including outcomes such as more engaged and useful interactions with health care providers and being better informed of the options available and their risks and benefits to support informed decision-making.

Acknowledgments
The authors would like to thank One Colorado, the Community Advisory Board, and the Integrated Transgender Clinic at the University of Colorado for their partnership and contributions to this project. They also want to thank University of Colorado Denver students Matthew Cirkovic and Ravisher Singh for their contributions as developers. The research reported in this publication was supported by the National Institutes of Health National Library of Medicine under award 5G08LM013200-02. The content is solely the responsibility of the authors and does not necessarily reflect the official views of the National Institutes of Health or the National Library of Medicine.

Authors' Contributions
All authors made substantial contributions to the methodological design and implementation, acquisition of data, and analysis and interpretation of the data and made critical revisions to the manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Transgender health information resources.
[PDF File (Adobe PDF File), 49 KB - humanfactors_v10i1e42382_app1.pdf ]

Multimedia Appendix 2
Transgender health information resources app credibility resources.
[PDF File (Adobe PDF File), 36 KB - humanfactors_v10i1e42382_app2.pdf ]

Multimedia Appendix 3
Transgender health information resources health information content.
[XLSX File (Microsoft Excel File), 47 KB - humanfactors_v10i1e42382_app3.xlsx ]

References


57. One Colorado homepage. One Colorado. URL: https://one-colorado.org/ [accessed 2021-09-27]


59. Zoom homepage. Zoom. URL: https://zoom.us/ [accessed 2021-07-09]

60. Zoom homepage. Zoom. URL: https://zoom.us/ [accessed 2021-07-09]


69. Trust it or trash it? Trust or Trash. URL: http://www.trustortrash.org/ [accessed 2022-05-16]


Abbreviations

CAB: Community Advisory Board
FG: focus group
HIT: health IT
LGBTQ: lesbian, gay, bisexual, transgender, queer or questioning
MARS: Mobile App Rating Scale
mHealth: mobile health
REDCap: Research Electronic Data Capture
SUS: System Usability Scale
TGD: transgender and gender diverse
TGHIR: transgender health information resource
uMARS: user version of the Mobile App Rating Scale

©Brad Morse, Andrey Soares, Bethany M Kwan, Marvyn Allen, Rita S Lee, Kristen Desanto, Brooke Dorsey Holliman, Kate Ytell, Lisa M Schilling. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 15.06.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Toward the Design of Sensing-Based Medication Adherence Aids That Support Individualized Activities of Daily Living: Survey and Interviews With Patients and Providers

Jacob T Biehl¹, BSc, MSc, PhD; Ravi Patel², BSc, PharmD; Adam J Lee¹, BSc, MSc, PhD

¹School of Computing and Information, University of Pittsburgh, Pittsburgh, PA, United States
²School of Pharmacy, University of Pittsburgh, Pittsburgh, PA, United States

Abstract

Background: Nearly half of Americans taking prescription medications do not take them properly. The resulting implications have a broad impact. Nonadhering patients develop worsened medical conditions and increased comorbidity of disease or die.

Objective: Clinical studies have shown that the most effective strategies for addressing adherence are those that are individualized to the context that each patient and situation require. However, existing aids for adherence are relatively ridged and poorly support adaptation to individual behaviors and lifestyles. The aim of our study was to better understand this design tension.

Methods: A series of 3 qualitative studies was conducted: a web-based survey of 200 Americans that investigated existing adherence strategies and behaviors and perception of how hypothetical in-home tracking technologies would assist adherence; in-person semistructured interviews with 20 medication takers from Pittsburgh, PA, that investigated personal adherence behaviors, which included demonstration of medication locations and routines as well as an assessment of hypothetical technologies; and semistructured interviews with 6 pharmacists and 3 family physicians to gain a provider perspective on patient adherence strategies, which included feedback on hypothetical technologies in the context of their patient populations. Inductive thematic coding of all interview data was performed. Studies were conducted consecutively, with the results informing the subsequent studies.

Results: Synthesized, the studies identified key medication adherence behaviors amenable to technological interventions, distilled important home-sensing literacy considerations, and detailed critical privacy considerations. Specifically, 4 key insights were obtained: medication routines are heavily influenced and adapted by and through the physical location and placement of medications relative to activities of daily living, routines are chosen to be inconspicuous to maintain privacy, the value of provider-involved routines is motivated by a desire to build trust in shared decision-making, and the introduction of new technologies can create further burden on patients and providers.

Conclusions: There is considerable potential to improve individual medication adherence by creating behavior-focused interventions that leverage emerging artificial intelligence (AI), machine learning (ML), and in-home Internet of Things (IoT) sensing technologies. However, success will be dependent on the technology’s ability to learn effectively and accurately from individual behaviors, needs, and routines and tailor interventions accordingly. Patient routines and attitudes toward adherence will likely affect the use of proactive (eg, AI-assistant routine modification) versus reactive (eg, notification of associated behaviors with missed dosages) intervention strategies. Successful technological interventions must support the detection and tracking of patient routines that can adjust to variations in patient location, schedule, independence, and habituation.

(JMIR Hum Factors 2023;10:e40173) doi:10.2196/40173

KEYWORDS
sensing; medication adherence; active intervention; self-management; patient care; medication; qualitative study; successful intervention; patient support
Introduction

Despite the life-saving and life-preserving power of medications, it is estimated that nearly 50% of patients do not properly adhere to their medications [1]. The World Health Organization (WHO) has described nonadherence as a “worldwide problem of striking magnitude” [1]. Nonadherence leads to “worsening condition, increased comorbid diseases, increased health care costs, and death” [2]. Although adherent behaviors vary across pharmacotherapies and diagnoses [3], they are broadly defined by the US Food and Drug Administration (FDA) as “getting prescriptions filled, remembering to take medication on time, and understanding the directions as prescribed” [4].

Nonadherence behaviors are costly to national health care infrastructures and capacities. In the United States, US $100 billion in additional hospitalization costs and US $2000 per patient in physician visits per year are attributed to nonadherence [5]. These costs manifest across a wide variety of pharmacotherapy situations. For instance, in an acute situation of postoperative coronary artery bypass graft surgery, patients are placed on β-blockers and other medications to reduce occurrences of atrial fibrillation [6,7]. In the United States, adherence to these important postoperative medications is estimated to be <55% [8] and contributes to a postoperative readmission rate of 26.7% [9]. Similar situations occur with chronic conditions, too. Recent studies show that only 78.7% of insulin-dependent individuals with diabetes are adherent to ≥80% of their injections, increasing risk of stroke and other cardiac complications [10,11]; just 82% of patients who have undergone a kidney transplant are adherent to immunosuppressants that are critical to prevent organ rejection [12], and just 25% of patients prescribed pre-exposure prophylaxis (PrEP) are adherent to >90% of doses, increasing risk of contracting HIV [13].

Despite the scale of the problem, research and development efforts to build effective technology aids and interventions to improve medication adherence are quite small compared with the size of the consumer health care technology market. In part, this limitation is governed by the multitude of policy and economic factors that influence adherence, especially with access to medications. However, there have been evident studies that have shown that technology-based monitoring and reminders can provide sustainable improvement in medication literacy and adherence behaviors across a broad population of users [14-16]. However, these studies have also noted the highly individual situations and sentiments that define and motivate medication-related behaviors, noting that technology approaches cannot have ridged approaches to adherence [17,18]. These cited clinical studies have shown that utility is greatest in interventions that are highly customized to address a patient’s specific adherence barriers.

Poor efficacy and the use of common metrics or techniques across detection, classification, and intervention are consistent themes in previous efforts. For instance, Chung et al [19] observed that health-tracking applications fail because of the use of one-size-fits-all metrics rather than individualized health goals. Similarly, Clawson et al [20] found that health-tracking technologies were often abandoned when users failed to realize the benefits of the technologies, in part because of the failure of the technologies to support specifics of lifestyles and expectations. They continue to argue that health-tracking technologies must work in users’ complex social lives and highly individualistic activities of daily living. Complementing the challenges of technology are the limited insights from clinical perspectives to address adherence as a behavior. Clinical interventions include limited data from the health system and self-report [21]. Clinical practices are limited in their ability to understand, interpret, or adapt to the implementation of technology to support adherence.

The lessons of past technical and clinical work must guide future health technologies’ design and development, specifically the need to use methods that seek solutions that easily integrate and support highly specialized, personal behaviors. However, understanding the path to achieving this goal is still an active design and technology exploration. To this end, this paper presents a set of in-depth formative studies to evaluate the potential for new and novel medication adherence interventions across a variety of clinical needs, lifestyles, attitudes, and behaviors concerning medication and overall health. The results of our studies suggest that the needs of users are extraordinarily diverse, influenced by a multitude of factors, including daily schedules, number and type of medications, and level of overall health. More so, though, complex sociotechnical factors also influenced needs and perceptions of technology. The adoption of reminder technologies, for instance, was heavily tied to the perception of self-independence. Technologies that incorporate automation and behavior modeling raise concerns not just about privacy but also about the judgment of personal health behaviors. The work described in this paper contributes a series of descriptive insights and associated implications for the design of medication adherence tools.

Methods

Overview
A series of three qualitative studies was conducted as follows: (1) a web-based survey that investigated existing adherence strategies and behaviors and perception of how hypothetical in-home tracking technologies would assist adherence; (2) in-person semistructured interviews with medication takers from Pittsburgh, PA, that investigated personal adherence behaviors, which included demonstration of medication locations and routines as well as an assessment of hypothetical technologies; and (3) semistructured interviews with pharmacists and family physicians to gain a provider perspective on patient adherence strategies, which included feedback on hypothetical technologies in the context of their patient populations. Studies were conducted consecutively, with the results informing the subsequent studies. In each of the following subsections, we describe each study design and the evaluation methodologies used.

All 3 studies were designed to further understand how existing artificial intelligence (AI), machine learning (ML), and Internet of Things (IoT) technologies could be used to track, learn, and inform on adherence behaviors and potential interventions. Most
prior work has focused on gaining population-level or intervention-level insights, which provide few practical design insights at the motivation and management levels. The methods of the 3 studies conducted in this work provide particularly useful investigations into how technologies could learn and leverage everyday routines and behaviors to drive interventions in adherence practice. These methods attempt to connect routines and behaviors with technology perception and acceptability.

**Formative Survey**

We surveyed 200 people, all living in the United States, using a web-based 52-question survey (Multimedia Appendix 1). The questions were a mix of multiple-choice, short-answer, and free-form questions. It was deployed using Amazon’s Mechanical Turk (Amazon Web Services) and conducted over a 2-week period in late September 2019 [22]. All respondents were screened before completing the survey to be currently taking at least one prescription medication. Tables 1 and 2 report the number of medications per respondent and the age of the respondents, respectively. There were no interaction effects between the number of medications and age, although the number of medications trended upward with age. Respondents were most likely to be under the care of a single physician (105/200, 52.5%), although many were under the care of 2 (61/200, 30.5%) or 3 (20/200, 10%). The number of prescribing physicians (i.e., respondents had an active prescription from this physician) was similar; most had only 1 (132/200, 66%), followed by 2 (44/200, 22%) and 3 (6/200, 3%). Each respondent was paid US $4 for completing the survey. The mean completion time was 9 minutes, 42 seconds (SD 5 minutes, 31 seconds).

**Table 1. Number of medications taken by survey respondents (N=200).**

<table>
<thead>
<tr>
<th>Number of medications taken</th>
<th>Respondents, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>71 (35.5)</td>
</tr>
<tr>
<td>2</td>
<td>54 (27)</td>
</tr>
<tr>
<td>3</td>
<td>36 (18)</td>
</tr>
<tr>
<td>4</td>
<td>15 (7.5)</td>
</tr>
<tr>
<td>5</td>
<td>12 (6)</td>
</tr>
<tr>
<td>≥6</td>
<td>12 (6)</td>
</tr>
</tbody>
</table>

**Table 2. Age of survey respondents (N=200).**

<table>
<thead>
<tr>
<th>Age range (years)</th>
<th>Respondents, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>18-24</td>
<td>6 (3)</td>
</tr>
<tr>
<td>25-34</td>
<td>55 (27.5)</td>
</tr>
<tr>
<td>35-44</td>
<td>61 (30.5)</td>
</tr>
<tr>
<td>45-54</td>
<td>35 (17.5)</td>
</tr>
<tr>
<td>55-64</td>
<td>25 (12.5)</td>
</tr>
<tr>
<td>≥65</td>
<td>18 (9)</td>
</tr>
</tbody>
</table>

**In-Home Interviews With Medication Takers**

A combination of in-person and remote video conference semistructured interviews was conducted with 20 medication-taking participants between February 2020 and September 2020. A total of 50% (10/20) of the interviews were conducted in person, and 50% (10/20) were conducted using videoconference. The in-home interviews were recorded to capture audio and video. This method also allowed for the observation of described adherence behavior within the context of broader home activities, medication location within the physical home layout, and second-party (e.g., spouse) participation in routines. While in the middle of conducting our interviews, the COVID-19 pandemic prevented in-person human participant studies at our institution; thus, 50% (10/20) of the participants were interviewed using a Health Insurance Portability and Accountability Act of 1996 (HIPAA)–compliant videoconferencing software. The sample size was based on previous health behavior studies and broader guidance across the discipline [23].

The interviews were conducted using a semistructured protocol. Guiding questions were used to explore interviewees’ perceptions of health status, current disease states and medications, medication adherence and daily personal routine, use of medication adherence tools, use of digital tools to support health and routine use in the home, and concerns about using technology to track and manage health. Participants were also asked to either draw a diagram of their home or describe their residence and describe the relationship between the participant and any other cohabitants and how, if at all, any cohabitants assisted in medication adherence or other experiences in health care. These guiding questions were followed by probing questions based on participants’ responses. The interviews ended with a series of structured feedback responses. Participants were provided with scenarios that described the same hypothetical technologies used in the formative survey. Interviewees provided
we recruited patients with hypertension as medication takers, promote adherence and the limitations of those approaches. As the mitigating context, and exploring existing methods used to provider prospective of adherence obstacles and assessment of clients’ medication adherence behaviors and compliance, focused on 3 main points of inquiry: overall concern with A semistructured interview protocol was used. The interviews expert interviews [24]. and community pharmacists. The sample size was based on organizations, including family practices, clinic pharmacists, authors’ professional networks. Participants were from several recruited providers through snowball sampling within the to interview providers of the medication takers interviewed. We interviewing providers and associated clients, we did not seek and August 2021. Owing to the institutional regulations of Health care provider interviews were conducted after the Pittsburgh region, with ages ranging from 27 to 71 (mean 47.75, SD 15.35; median 43) years. A total of 70% (14/20) identified as female, and 30% (6/20) identified as male. Participants reported the total number of prescription medications taken; the number ranged from 2 to 13 (mean 5.79, SD 3.65; median 5). The interviews lasted between 90 and 120 minutes. Each participant was compensated with US $40 for taking part in the study.

Interview recordings were outlined using the interviewers’ notes and transcribed to identify quotes. Interview transcriptions and notes were analyzed using a general inductive approach by 4 researchers. General topics were coded (eg, location in which the medications were kept, current medical conditions, and relationship with cohabitants). Cross-coding was performed on a random sample of 30% (6/20) of the interviews. Once coded, the same 4 investigators engaged in a systematic review of the coded interviews to organize them into categorical findings.

**Interviews With Health Care Providers**

Health care provider interviews were conducted after the medication taker interviews were completed. These were conducted at the location of practice (retail pharmacy, medical practice, or hospital) or via teleconference to accommodate the provider’s comfort with in-person interaction during the COVID-19 pandemic. A total of 9 interviews were conducted: 6 (67%) with pharmacists and 3 (33%) with family practice physicians. The interviews were conducted between July 2021 and August 2021. Owing to the institutional regulations of interviewing providers and associated clients, we did not seek to interview providers of the medication takers interviewed. We recruited providers through snowball sampling within the authors’ professional networks. Participants were from several organizations, including family practices, clinic pharmacists, and community pharmacists. The sample size was based on broader guidance across the discipline [23] and guidelines for expert interviews [24].

A semistructured interview protocol was used. The interviews focused on 3 main points of inquiry: overall concern with clients’ medication adherence behaviors and compliance, provider prospective of adherence obstacles and assessment of the mitigating context, and exploring existing methods used to promote adherence and the limitations of those approaches. As we recruited patients with hypertension as medication takers, we asked providers to answer questions specifically about patients with hypertension, although they would often indicate that responses applied more broadly to their client populations. Providers were presented with scenarios that described hypothetical technologies. They were asked to describe how useful they felt these technologies were overall and provide insights into the type of clients they felt would be more and less receptive to using them. We did not collect the age of the providers; all had >10 years of practice in their roles. The interviews lasted approximately 60 minutes. Participants were compensated with food or a gift card for lunch at a local restaurant. We used the same analysis techniques and cross-coding procedures used for the medication taker interview data. The full semistructured interview protocol is provided in Multimedia Appendix 3.

**Ethics Approval**

Each study was reviewed and approved by the University of Pittsburgh institutional review board (STUDY19080322, with protocol title “Technology for Prescription Adherence”).

**Results**

The 3 studies were performed sequentially; each study was informed by the previous one. Thus, we present the results sequentially.

**Findings From the Formative Survey**

The formative survey provided insights into broad adherence practices, existing technology use, and sentiment on potential technology aids. Common behaviors associated with missed dosages were indicated and rated for frequency (daily, weekly, monthly, and yearly). The reasons were presented as a list of behaviors derived from a synthesis of previous studies. Responses are summarized in Table 3. A Pearson contingency analysis indicates that a significant difference was observed across groupings by reason ($\chi^2_42=419.1; P<.001$).

Respondents were asked to rate their use of commonly recommended and prescribed adherence aids, namely, pill box or medication organizer and diary or calendar. Indications and frequency of use were also measured. As summarized in Table 4, the means trended positively with frequency of use. A Kruskal-Wallis test on differences in the effectiveness ratings (grouped by use) was significant for both the pill box ($\chi^2_3=60.5; P<.001$) and diary ($\chi^2_5=15.0; P<.002$) aid types. Only 14% (28/200) of the respondents indicated use of an existing medication-tracking digital application. Ranking the top features, of these 28 respondents, 14 (50%) indicated reminders and notifications, 6 (21%) indicated the ability to track medication doses, and 4 (14%) indicated signaling the pharmacy to refill the prescription. Of the 172 respondents not using an application, 61 (35.5%) responded that the use of a smartphone for tracking medications was too cumbersome, 47 (27.3%) expressed concerns about privacy and security, and 27 (15.7%) indicated that they did not use their smartphone in the same location in which they stored and took their medications. Respondents stored medications in a variety of locations: the bedroom (80/200, 40%), the kitchen (74/200, 37%), and the bathroom (80/200, 40%).
(41/200, 20.5%). A total of 16% (32/200) carried medications on their person. Many (50/200, 25%) indicated that they kept medications in more than one location in their home, but only 1% (2/200) indicated keeping medications outside their home.

Respondents were given brief descriptions of 6 hypothetical smart home technologies. These technology descriptions were based on generic approximations of technologies that could be imagined based on current state-of-the-art IoT technologies. The specific descriptions are provided in Textbox 1. Respondents were asked to “rate the perceived utility of [the] technology.” Respondents were then asked to rate whether the technology would be “more or less useful than their current medication adherence practices.” Table 5 summarizes respondents’ feedback. The results confirmed that adherence remains difficult for a broad set of American adults, young and old. Across failures to adhere, unintentional reasons (eg, forgetting the dose and being away from the medication) were more frequent than intentional reasons (eg, avoiding side effects and having no symptoms). A common theme among respondents was that there were multiple reasons for missed doses, reflecting the complexity and deep embedding of adherence behaviors within daily activities and routines. The usefulness ratings of the hypothetical technologies strongly connected with perceived use in improving adherence. These results motivate the need for technologies that reinforce habituation, have low barriers to use, and adapt to the complexity of medication storage and interaction.

These results were void of the context necessary to understand how and why people choose their medication adherence routines, how they use their selected tools and aids to support those routines, and the factors surrounding the daily challenges of medication adherence. These in-depth questions necessitated and informed the design of a qualitative approach based on in-depth interviews with patients and health care providers. In study 2, we interviewed patients. In study 3, we interviewed health care providers.

Table 3. Respondent frequency of reasons for missing medication doses (N=200).

<table>
<thead>
<tr>
<th>Reason</th>
<th>Missed dose at least once per..., n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Year</td>
</tr>
<tr>
<td>Medication dose forgotten</td>
<td>136  (68)</td>
</tr>
<tr>
<td>Medication in a different location</td>
<td>106  (53)</td>
</tr>
<tr>
<td>Activity prevented taking medication</td>
<td>85  (42.5)</td>
</tr>
<tr>
<td>Medication supply ran out</td>
<td>51  (25.5)</td>
</tr>
<tr>
<td>Avoidance of side effects</td>
<td>47  (23.5)</td>
</tr>
<tr>
<td>No symptoms present</td>
<td>43  (21.5)</td>
</tr>
<tr>
<td>Making the medication last longer</td>
<td>43  (21.5)</td>
</tr>
<tr>
<td>Social situation made it inappropriate</td>
<td>34  (17)</td>
</tr>
</tbody>
</table>

Table 4. Perceived effectiveness of existing aid by frequency of use (N=200).

<table>
<thead>
<tr>
<th>Frequency of use</th>
<th>Pill box</th>
<th>Diary</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Participants, n (%)</td>
<td>Rating, mean (SD)</td>
</tr>
<tr>
<td>Always</td>
<td>44 (22)</td>
<td>4.66 (0.71)</td>
</tr>
<tr>
<td>Most of the time</td>
<td>23 (11.5)</td>
<td>4.17 (0.49)</td>
</tr>
<tr>
<td>Approximately half of the time</td>
<td>10 (5)</td>
<td>3.20 (0.79)</td>
</tr>
<tr>
<td>Sometimes</td>
<td>33 (16.5)</td>
<td>3.12 (0.82)</td>
</tr>
</tbody>
</table>

^aN/A: not applicable.
Textbox 1. Description of the hypothetical technologies used in the formative survey.

<table>
<thead>
<tr>
<th>Technology</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proximity Notifications (PN)</td>
<td>A technology that would detect and alert when you are near your medications, providing in-situ notifications for scheduled doses.</td>
</tr>
<tr>
<td>Proximity Notifications to Caregiver (PN-Care)</td>
<td>A technology that would detect and alert when a caregiver, family member, or trusted aide is near your medications.</td>
</tr>
<tr>
<td>Pre-Departure Reminder (PDR)</td>
<td>A technology that would detect and alert when you are about to leave your home without taking scheduled medications.</td>
</tr>
<tr>
<td>Routine Change Detection (RCS)</td>
<td>A technology that would learn more about your behaviors and movements when you are within and away from your home. It could use this information to suggest times and locations for taking medications that could lead to improved adherence.</td>
</tr>
<tr>
<td>Routine Change Detection Supported with Healthcare Providers (RCS-Team)</td>
<td>A technology that would learn more about your behaviors and movements when you are within and away from your home. Behaviors and movements would be summarized and made available to your healthcare professionals. These summaries could be used to improve medication selection, scheduling, dosing, and other instructions provided by your healthcare professionals to improve adherence.</td>
</tr>
<tr>
<td>Medication with Food (MwF)</td>
<td>A wearable or smart home technology that would learn more about your behaviors to classify when you are eating a meal. The technology could help remind you to take medications that need to be taken with a meal or simply help establish a routine of taking medications with meals.</td>
</tr>
</tbody>
</table>
Table 5. Perceived usefulness and impact on existing personal adherence for each hypothetical technology (N=200).

<table>
<thead>
<tr>
<th>Hypothetical technology and usefulness rating</th>
<th>Participants per usefulness rating, n (%)</th>
<th>Perceived impact on existing adherence regimen, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PN&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td>Much worse  Somewhat worse  Same  Somewhat better  Much better</td>
</tr>
<tr>
<td>Very</td>
<td>53 (26.5)</td>
<td>0 (0)      2 (3.8)     25 (47.2)   11 (20.1)   15 (28.3)</td>
</tr>
<tr>
<td>Somewhat</td>
<td>96 (48)</td>
<td>8 (8.3)    40 (61.7)   45 (46.9)   3 (3.1)    0 (0)</td>
</tr>
<tr>
<td>Not</td>
<td>51 (25.5)</td>
<td>22 (43.1)  24 (47.1)   5 (9.8)     0 (0)      0 (0)</td>
</tr>
<tr>
<td>PN-Care&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very</td>
<td>29 (14.5)</td>
<td>0 (0)      0 (0)       5 (17.2)    8 (27.6)    16 (55.2)</td>
</tr>
<tr>
<td>Somewhat</td>
<td>61 (30.5)</td>
<td>0 (0)      3 (4.9)     36 (59)     17 (27.9)   5 (8.1)</td>
</tr>
<tr>
<td>Not</td>
<td>110 (55)</td>
<td>32 (29.1)  26 (23.6)  49 (44.5)   1 (0.9)     2 (1.8)</td>
</tr>
<tr>
<td>PDR&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very</td>
<td>96 (48)</td>
<td>0 (0)      0 (0)       5 (5.2)     34 (35.8)   56 (38.9)</td>
</tr>
<tr>
<td>Somewhat</td>
<td>71 (35.5)</td>
<td>2 (2.8)    3 (4.2)     19 (26.8)   44 (62)     3 (4.2)</td>
</tr>
<tr>
<td>Not</td>
<td>34 (17)</td>
<td>11 (32.8)  5 (14.7)   17 (50)     1 (2.9)     0 (0)</td>
</tr>
<tr>
<td>RCS&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very</td>
<td>38 (19)</td>
<td>0 (0)      0 (0)       1 (2.6)     18 (47.4)   19 (50)</td>
</tr>
<tr>
<td>Somewhat</td>
<td>71 (35.5)</td>
<td>0 (0)      3 (3.4)     39 (44.3)   39 (44.3)   7 (8)</td>
</tr>
<tr>
<td>Not</td>
<td>74 (37)</td>
<td>27 (32.6)  16 (21.6)  31 (41.9)   0 (0)      0 (0)</td>
</tr>
<tr>
<td>RCS-Team&lt;sup&gt;e&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very</td>
<td>27 (13.5)</td>
<td>0 (0)      0 (0)       0 (0)       9 (33.3)    18 (66.6)</td>
</tr>
<tr>
<td>Somewhat</td>
<td>78 (39)</td>
<td>0 (0)      2 (2.5)     34 (44.3)   38 (48.7)   4 (5.1)</td>
</tr>
<tr>
<td>Not</td>
<td>95 (47.5)</td>
<td>34 (35.8)  24 (25.3)  35 (36.8)   2 (2.1)     0 (0)</td>
</tr>
<tr>
<td>MwF&lt;sup&gt;f&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very</td>
<td>60 (30)</td>
<td>0 (0)      0 (0)       3 (5)       23 (38.3)   34 (56.6)</td>
</tr>
<tr>
<td>Somewhat</td>
<td>85 (42.5)</td>
<td>0 (0)      1 (1.2)     34 (40)     46 (54.1)   4 (4.7)</td>
</tr>
<tr>
<td>Not</td>
<td>55 (27.5)</td>
<td>12 (21.8)  11 (20)    30 (54.5)   2 (3.6)     0 (0)</td>
</tr>
</tbody>
</table>

<sup>a</sup> PN: Proximity Notification.
<sup>b</sup> PN-Care: Proximity Notifications to Caregiver.
<sup>c</sup> PDR: Pre-Departure Reminder.
<sup>d</sup> RCS: Routine Change Detection.
<sup>e</sup> RCS-Team: Routine Change Detection Supported with Healthcare Providers.
<sup>f</sup> MwF: medication with food.

Findings From the In-Home Interviews With Medication Takers

Overview

Survey responses did not have adequate context to fully understand how the environment and living configurations of respondents affected their medication adherence routines, how they used their selected tools and aids to support those routines, and the factors surrounding the daily challenges of medication adherence. These in-depth interviews provided a deeper design understanding based on the broader understanding provided by the formative survey. The interviews provided considerable qualitative insights into the causes of missed dosages and the use of medication adherence aids, as well as usefulness and utility insights for the hypothetical technologies.

Location and Routine as an Adherence Aid

The interviews found the location and placement of medication to be a key component of medication adherence behavior. All interviewed medication takers (20/20, 100%) provided very detailed explanations and rationales for where medications were kept. The desire to position medications so that they were near and accessible during activities of daily living was the most substantial motivation (17/20, 85%). Most interviewees placed their medications in locations that were proximal to where...
Adherence Behaviors Are Private

Medication management behaviors, including adherence, were highly private activities for most interviewees. A total of 75% (15/20) of the participants lived with someone whom they felt was a “trusting relationship,” but only 27% (4/15) indicated that their cohabitant partner assisted in adherence-related activities. For instance, medication taker 19 noted that she used to keep her medications in a kitchen cabinet but once found her child standing over the sink holding the medications and noted that it was a “not going to happen anymore” moment.

Interestingly, the locations were highly individual to each person, with no single rationale emerging as a general trend. Participants recognized in many cases the idiosyncrasies of their rationale. For instance, medication taker 20 kept the bottle for his pills in an ottoman in his bedroom, noting when explaining that “I know this sounds weird, but it works for me.” Medication taker 1 kept her pills in an old cookie tin and carried it around with her from room to room in her home:

At one time, I was a scatter brain, my meds would be all over the place. “Where’s this? Where’s that?” I said, “I can’t live like this.” Everything has to be in order. When everything is scattered, I’m out of control.

Adherence Behaviors Are Private

Medication management behaviors, including adherence, were highly private activities for most interviewees. A total of 75% (15/20) of the participants lived with someone whom they felt was a “trusting relationship,” but only 27% (4/15) indicated that their cohabitant partner assisted in adherence-related behavior. This was similar to the number of participants (5/15, 33%) who indicated that their trusted partner could communicate with medical providers on their behalf. We found that this privacy was centered on concerns of independence and perceived burden. For instance, when medication taker 4 was asked if his wife helped with his adherence, he responded with “no, and I don’t wear Velcro shoes either,” insinuating that, in addition to managing his own medication adherence, he was also able to put his shoes on independently. Similarly, medication taker 14 stated:

There are certain things as a grown up that you do by yourself.

For those who did receive help from trusted partners or cohabitants, it was usually motivated by their medical condition having a dramatic impact on their ability to independently perform activities of daily living. For instance, medication taker 6 had an injury preventing him from leaving his bed. A few of the older medication takers interviewed also indicated an openness to receiving help when they could not manage adherence on their own. For instance, medication taker 11 stated the following:

If I needed help, I wouldn’t be afraid to ask, I just don’t need help yet.

The private nature of medication behaviors was also shown in interviewees’ concerns about interventions that involved other individuals. When asked about smart speaker reminders for medication, medication taker 17 indicated that she was hesitant to use her Google Home for personal health as “if [she] had friends over, [she] would not want to announce.” Independence concerns were also noted in the hypothetical technology ratings, with only 30% (6/20) of the interviewees indicating that they found Proximity Notifications to Caregiver (PN-Care) to be very useful, whereas most (12/20, 60%) found it not useful.

Role of Health Care Providers in Medication Adherence

Involving health care providers in medication adherence tools elicited mixed opinions. Many felt that it would be helpful in establishing trust by relaying information about adherence behaviors to physicians. In total, 55% (11/20) of the medication takers indicated that Routine Change Detection Supported with Healthcare Providers (RCS-Team) would be very useful, 25% (5/20) indicated that it would be somewhat useful, and 20% (4/20) indicated that it would be not useful. For those indicating usefulness, the motivation was primarily centered on establishing a stronger understanding of medication behavior and effectiveness with the provider. For instance, medication taker 14 stated the following:

I would like the doctor to have a more factual set of information for I’m doing as opposed to pre-conceived notions that they have.

Medication taker 15 stated that the hypothetical technology “would be very helpful if it would help with adjustments of something not known to self,” indicating that health care providers may use it to improve diagnosis or treatment.

Medication takers expressed concerns about the reach of technology; in particular, concerns about privacy and the negative impact on care were expressed. Medication taker 11 was very articulate in her concerns:

There’s a thin line there with how much your doctor needs to know. I guess it would depend at what you’re looking at. If it were to cure cancer, I’d say it’s very useful, but if it were for something minor, I wouldn’t be for it.
Medication takers felt that it was simply unnecessary for the physician-patient relationship. Medication taker 12 relayed the following:

*If I want to express something to my health care provider, I would just tell them, I don’t need an app to do it.*

**Role of Technology in Health**

Privacy concerns with the hypothetical technologies were mixed. Medication taker 1 was very split on achieving the benefits at the cost of privacy:

*I heard [about this] technology that learns your behaviors. I mean it’s scary to give technology that much power. I don’t know what to do. I’m for it to a certain degree. It’s good that it does learn your behavior, but it’s too much.*

Medication taker 12 provided a more succinct value proposition:

*Security is always a good thing to keep in mind, I would really have to want a technology to get it.*

Other medication takers expressed less concern, focusing on the perceived low utility of their personal health information to others. Medication taker 11 stated the following:

*I don’t worry too much about my privacy. I have a mindset that I don’t have that much to hide, if you want to know I don’t care.*

Medication taker 8 echoed a similar sentiment—“To be honest, I’m an open book. Information like that doesn’t bother me at all”—referring to personal medical history. Overall, we found the lack of privacy concerns to be a somewhat stark contrast to the private nature of medication adherence.

**Findings From Interviews With Health Care Providers**

The provider interviews were an important complement to the medication taker interviews. To allow for a straightforward comparison, we organized the results in the same structure as the findings from the medication taker interviews.

**Location and Routine as an Adherence Aid**

The effectiveness of routines and aids was more sobering from the perspective of pharmacists and physicians. In our interviews, we quoted the literature that indicated that <50% of patients overall are adherent; when asked if they agreed, all providers (9/9, 100%) responded with overwhelming agreement. When asked to describe the reasons for nonadherence, physician 2 noted that most nonadherence behaviors are not intentional:

*In my [patients] it tends to be more ability to pick up medicines, ability to be organized enough to manage the medicines, that sort of thing. More so than their overt, “I’m not going to take this.”*

Physician 3 noted similar reasons and was empathetic to the patients:

*I mean, I think they all try to...it’s that they forget they fall asleep at night, so they forget their nighttime dosage. For most people, I think they want to be compliant. It is a matter of when you don’t have a symptom. How do you remember to take medication every day?*

The most common mechanism providers have to monitor adherence is refill statistics, which are commonly available to both pharmacists and primary care physicians. They noted that refill cadence often does not match the prescription, indicating that doses have been missed. Providers often discuss this with clients, as pharmacist 2 stated:

*I go through the normal questions, are you having side effects? Let’s talk to your doctor about lowering doses. There’s a reason you’re not taking it. Are you having side effects? [Do you] just forget to take it?*

Physician 1 provided a perspective that emphasized the prevalence of nonadherence but also noted the current limitations in their ability to understand client adherence behaviors:

*I think that their compliance, because I have written an order; they say that they’re engaged. [But] we don’t really understand the level of non-compliance. And, therefore, to put that [effort] into [an] understand why this [happens] is hard.*

This was echoed by pharmacist 4:

*I definitely [try] to give them benefit of the doubt, but I feel like even after I go through all of those questions and the patient’s like no I take every day, no problems, no issues. I think if they could just be honest with us and help us, to help us identify those adherence issues and we can help them solve it.*

Prepackaged medications received mixed reviews from the health care providers interviewed. Providers indicated that this approach works well for patients who lack home health assistance either from a family member or outside organization. Providers also pointed out a considerable limitation: a prepackaged medication routine can quickly become confusing and difficult for patients to manage when medications or dosages are changed. Physician 2 stated the following:

*...anytime you make an adjustment, it’s a delay in them getting the new meds so and then it’s sort of a rigmarole to get them changed. And I’ll have patients that ended up with other meds that are now in separate bottles. And so, somebody has to be orchestrating the pill packaging.*

Providers were supportive of pill organizers, and they indicated that they often recommended that patients use them. Physician 2 noted that she had pill organizers at her clinic and would, on occasion, even demonstrate to patients how to fill them. Interestingly, providers often noted that using an organizer alone is not sufficient; providers stated that it also matters where the pill organizer is kept. Physician 3 recalled a recent conversation about the location of the pill organizer:

*He has [his medications] beside his bedside table. But he falls asleep and doesn’t remember.*

Providers were aware of the diversity of behaviors regarding where their clients kept their medication. Physician 3 noted that she would often review the location of her clients’ medications...
as part of adherence discussions. Providers indicated that they often had to review medication storage locations as living situations change and can disrupt adherence routines. For instance, physician 1 recalled several conversations with his low-adherence patients and noted the following:

> It's one of the things we find often is the medications in the bedroom, or it's in the living room, and [they say they] spend all the time in the bedroom in the summer, because that's for the air conditioner, or something.

**Adherence Behaviors Are Private**

Privacy and independence concerns between a medication taker and a provider rarely exist. This mirrors the results from the patient interviews and is perhaps not surprising considering that physicians and pharmacists are trusted individuals who prescribe and dispense the medications. Providers were asked to reflect on how their clients rely on family members and cohabitants for adherence. The responses were consistent with those of medication takers. Family and friends were used for transportation to the physician’s office and pharmacy or were asked to pick up refills. Physicians also noted that it was very rare for clients to have others present in the examination room. Physician 2 noted the following:

> I would say probably about 5 or 10% of my patients come with somebody.

**Role of Health Care Providers in Medication Adherence**

Providers expressed great interest in having tools that could help them understand client medication adherence behavior. Pharmacist 1 expressed interest in getting a sense of overall adherence trends, noting the following:

> ...if [patients] reported on a weekly basis of like how many days that they missed, or what days that they missed their medications, it would be super helpful.

Physician 1 went further, indicating that more information would have actionable value, stating the following:

> ...system, and it says, you know, [patient name] took his medicine two days out of 10 this week, or 10 times in the last three days, that's very helpful information. Because if the system allows for health care proxy kind of function, then that's, that's a level of sort of monitoring compliance, that could be very helpful. That doesn't involve us or could escalate, you know, hey, [patient name] is really not taking this medicine. What else can we do here? Can we change the system?

In many ways, this was similar to comments from medication takers, who wanted to use data to force action or attention from a care provider. Similarly, providers also mentioned concerns about privacy. For instance, when physician 2 was asked for feedback on RCS-Team, she stated the following:

> It seems like a cool idea. But sort of invasive [of] patients with privacy issues, it seems like big brother's watching, I don’t like it.

**Role of Technology in Health**

Provider feedback on technology opportunities was overly focused on the limited time they had with patients and worries that introducing a new technology would take time away from other important tasks. When asked about reviewing patient data in the context of RCS-Team, physician 2 stated the following:

> ...would I have time? My immediate answer is no, of course not, I don’t have time to do what I need to do already. What am I skipping to add that in? There is no time to do anything extra. Unless somehow miraculously, it helps with something else so that it slims down, you know, tears down the rest of the junk I have to do.

In addition, providers were concerned about the burden they would feel to make the data actionable. Physician 1 expressed an articulate vision of the long-term insights that technology could provide into patient health but struggled with how he could expect that exploring data and formulating insights into a useful recommendation to patients would fit into his busy work practice:

> [to know] their blood pressure’s normal 85% of the time and they’re taking their medicine 87% of the time and [the medication is kept] at the front door, often, maybe that part could get us to 90% of the time, like, you know, something that would sort of look at this a little intelligently and say, so the problems not necessarily compliance, but with that 13% compliance. It just needs to tell me where the opportunities are.

**Discussion**

**Principal Findings**

These studies indicated several key insights to better leverage technologies to support challenges with medication adherence. Similar to the adherence challenge itself, there were many complicated sociotechnical concerns raised that will challenge the design of these future technologies. These range broadly from the diversity of each individual’s adherence behaviors and routines to the complications of designing technology that can be flexible to support a diverse set of needs at a low cost. In particular, the analysis uncovered the need for technologies that do not just target one behavior or action but are broad and flexible enough to achieve sustained utility as needs and practices shift.

**Strengths and Limitations**

This study contributes an understanding of the perceived use of technology-aided medication adherence tools and interventions. Although current approaches to medication adherence can implement static routines or idealized behavior, this study can inform the development of sensing, persuasive, and other technologies that meaningfully leverage users’ current behaviors and adapt to changing needs. The most substantial limitation is that the collected user sentiments were based on previous behavior, preference, and perception of hypothetical technologies, not on actual use. The formative survey was limited to 1 Western-culture country, and the interviews were
focused on a specific geographic region of that country. Future investigations are needed to understand how cultures, access to health care, governmental influence, and many other factors affect adherence behaviors and adherence technology aids. Our investigation also focused on medication takers, pharmacists, and primary care physicians. It did not include other important entities such as health care system administrators or operators, insurers, or government aid agencies. Future investigations with these other stakeholders will be needed to fully translate design insights into actionable technologies for changes in practice and outcomes [25].

**Comparison With Prior Work**

Most medication adherence studies have focused not on intervention assessment but on the clinical outcomes of adherence within particular populations and diagnoses [26]. These studies can be epidemiological in focus or more targeted in scope, for instance, clinical trials of new medications where adherence is a requisite for measuring efficacy. Across studies, methods for collecting adherence data are surprisingly manual and analog. Such studies, their methods, and their results do not provide deep insights into adherence behaviors.

Distilling poor adherence to a specific behavior or factor is not possible [27]. Unintentional patient behaviors are understood to be a substantial contributor to nonadherence [17]. For instance, a recent study on patients with heart failure found that nearly 50% of missed doses were attributed to forgetfulness [28], matching the results from the studies presented in this paper. Interventions to address unintentional nonadherence have not had great success. A large study comprising 53,480 patients and 2 years of prescription data showed that low-cost reminder devices did not improve adherence [29]. Analyses of interventions have found that “to improve adherence effectively, there is a need for a tailored approach based on the type and cause of nonadherence and the specific needs of the patient” [17]. Recent surveys of adherence studies further support the need for tailored aids that allow for effective counseling and feedback, both automated and provider generated [30,31]. Bateman [32], reflecting on study adherence aids for patients with asthma, noted the need for “customized patient-friendly treatment that anticipates and accommodates usual behavior…is more likely to achieve the desired goal of disease control.” We believe that our studies provide new design insights toward meeting this goal.

Existing technology-based adherence intervention research has established that personalized interventions are essential for technologies to be efficacious [14,33]. Previous work has also illustrated that this level of personalization is difficult using traditional technologies (eg, smartphone apps) [14]. Other work has established the need for technologies to leverage known behaviors around routines, especially those specific to space use and medication storage [34-38]. The studies by Palen and Aaløkke [33] and by Dalgaard et al [39] have shown the importance of technologies that involve care providers, particularly in relation to patient-provider interactions. Our work builds on these past works; we advanced the design understanding of how technologies can be used to drive personalized interventions that leverage a combination of behavior sensing and models of effective adherence and self-monitoring.

**Implications for Practice, Research, and Design**

**Overview**

These 3 studies provided several key insights to better leverage technologies to support challenges with medication adherence. Similar to the adherence challenge itself, there were many complicated sociotechnical concerns raised that will challenge the design of these future technologies. These range broadly from the diversity of each individual’s adherence behaviors and routines to the complications of designing technology that can be flexible to support a diverse set of needs at a low cost. In particular, the analysis uncovered the need for technologies that are broad and flexible enough to achieve sustained utility as needs and practices shift.

**Routine-Aware Aids**

One of the most striking adherence behaviors uncovered was the extensive organizational routines that punctuate activities of daily living, for instance, placing medicine on nightstands to be the first point of attention in the morning and flipping medicine bottles over once taken in the morning. Most of these behaviors were adopted because they allowed the repetitive, routine nature of daily medication activities to be, as explained by medication taker interviewees, difficult to ignore, skip, or alter. Further, we found that pharmacists and primary care physicians encourage their clients to adopt these behaviors and even spend time with patients to discuss and strategize these efforts. In contrast, our survey found very few people who used technology reminders or alarms to support medication adherence routines; in fact, our interviews found only 1 medication taker who used an alarm on their smartphone. Alarms, along with existing smartphone app–based medication adherence tools, lack activity context. Specifically, alarms and apps will present reminders when medications should be taken, not when and where they should be taken. As a medication taker interviewee stated, an alarm going off when one is away from one’s medication is a “useless” reminder; it would simply be dismissed as taking the medication would not be possible or convenient at that moment. This implication raises serious concerns about the sustainability of adherence behavior change interventions documented in the adherence randomized clinical trial literature.

In contrast to the lack of perceived usefulness of alarms, the presented hypothetical Proximity Notification (PN) was perceived as very or somewhat useful (150/200, 75% in the survey and 13/20, 65% in the interview). We believe that this large difference is a result of the proximity-based notification’s ability to capture the where component of medication adherence behaviors. Furthermore, in our formative survey, we asked respondents to rate the usefulness and perceived impact on medication adherence behaviors. The ratings were highly related: the higher the usefulness, the higher the perceived impact. On the surface, this relationship is not surprising, but it explains a broader expectation that potential users of technologies need to understand and appreciate the impact on their behaviors and routines. Indeed, when speaking with providers, patients who were able to make this connection in
their behaviors were the most successful with adherence. In fact, many health maintenance routines have similar when and where routine-driven use, which suggests that design for this time or place context needs to be more deeply considered in the development of future technology aids.

**Aids Not Just for “Every Day” but for All Days**

Previous research has shown that technology adoption and use are messy in the real world. Clawson et al [20] concluded that self-monitoring health technologies need to be designed to accommodate the “ever changing dynamics of individuals’ lives.” Our investigations suggest that there is a need to address dynamics that are not changing over time but that are a part of the regular messiness of activities of daily living. Unsurprisingly, we found that routines vary from one day to the next. One day might be a trip to church, another might involve volunteering at the community center, and yet another might involve sports activities. These daily dynamics wreak havoc on a person’s ability to follow a tight daily routine, punctuated by small repetitive behaviors such as taking medications. Further, providers noted that changes in dosages or additions of medications were danger zones for nonadherence as they involved a routine change. Particularly troublesome were the times of transition in care (eg, being discharged from inpatient care).

Future technologies for health interventions will need to use adaptive aids to accommodate the dynamics in these activities of daily living or will need to be agnostic to the daily dynamics. For instance, medication takers in our studies suggested that tools that can react to these changes were perceived as useful. Specifically, ratings on Pre-Departure Reminders (PDRs) were high (166/200, 83% in the surveys and 13/20, 65% in the interviews), as were the ratings on perceived impact (Table 5). Similarly, reminders centered on mealtime also had considerable usefulness ratings (146/200, 73% in the surveys and 13/20, 65% in the interviews) and potential to affect adherence behaviors. Managing these types of dynamic aids, particularly how they can adapt to day-to-day and week-to-week variations in a scalable manner, engage with individuals and their personal context and circumstances, and effect positive changes via personalized and context-driven interventions, remains a substantial challenge.

**Reactive Versus Proactive Aids**

Our combined studies provided an interesting context to understand the broader expectations of intervention functionality for health-focused technologies. Although it was not surprising to confirm differing opinions from potential users (and even differences in opinions among the care providers), it was interesting to observe a divide between medication takers who preferred aids that were reactive in their design (eg, the PN and PDR notification hypothetical technologies) and those who preferred aids that were proactive in their design (eg, the Routine Change Detection [RCS] and RCS-Team hypothetical technologies). When asking medication takers to contextualize their preferences, it became apparent that there were fundamental differences in the motivation for these preferences. In particular, reactive aids were strongly preferred by medication takers who preferred technologies to, as a medication taker said, “help me be me.” These were medication takers who had clear, well-defined daily medication routines; had high personal motivation to maintain or improve their health; and (most importantly) did not feel the need for a technology recommending or forcing the adoption of new routines or changes to daily activities. In contrast, we had medication takers who clearly wanted the technologies to “help me be a better me” (adaptation of the previous quote). These interviewees were less organized in their routines or simply overwhelmed by their medication regimens or activities of daily living. Simply put, they were looking for aids that could provide answers and insights to make their lives better.

Indeed, medication takers wanted the technologies’ proactiveness to extend to health care providers, initiating changes in dosage or medication selection. Providers were also interested in leveraging these technologies; however, they were greatly hesitant to embrace any technology that would be an additional responsibility (and time commitment) for their professional practice. Future technologies must be designed to recognize the preferences and needs of the individual medication taker. Ideally, technologies should also create pathways to bridge their users from one intervention strategy to another. The technologies should bring more independence and confidence to those who can begin to master their personal health and provide increasing support to those who struggle, perhaps through strategic data sharing and the involvement of a medication taker’s pharmacist and primary care physician.

**Aids to Promote Independence**

A common sentiment of concern expressed by patients was the perception of their independence in their own care. We found a strong aversion to enabling other people to participate in medication adherence activities and behaviors. Even in the case of medication takers who lived with cohabitants, most of whom were also intimate partners, they were nonetheless not included in activities related to their personal health. Pregnancy and debilitating conditions were the only exceptions, and still, independence was guarded. These sentiments carried over to technology perception. Our lowest-rated hypothetical technology was one that explicitly involved other cohabitants in medication adherence behaviors. Further, a concern expressed regarding smart speakers such as Amazon’s Alexa was that they would share these private behaviors with the household. This leads to an important design lesson: health technology aids must allow users to feel and be in control, assuring them that they are independent and not dependent on others or even the technology itself. It is important to note that autonomy is not just a design goal but also a clinical goal [40]. Aids that involve or promote engagement with health care providers can improve overall health care utility and use. Technologies that can represent a strong connection between patients and health care providers can capitalize on this underrealized potential. As evident in our conversations with providers, there was a strong desire for technologies that could empower patients to better understand and take control of their health care.

**Conclusions**

Adherence to prescription medications is a ubiquitous and nuanced challenge for many people. Technology offers a
promising contribution to this challenge based on its ability to learn from individual behaviors, needs, and routines and tailor interventions accordingly. Our formative survey results reaffirmed the challenges that people face when attempting to be adherent to their medications and characterized the relationship between technology and users’ needs, motivations for use, and expectations. Our interviews with medication takers and health care providers add a rich context to define both challenges and opportunities for technology. To address user needs and expectations, technology must support routines that vary in style of intervention, level of independence, and ability to inspire habituation. The findings also showed that technology will provide the most value when it is able to adapt to unanticipated changes in these variables.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Survey questions used in study 1.[PDF File (Adobe PDF File), 196 KB - humanfactors_v10i1e40173_app1.pdf]

Multimedia Appendix 2
Semistructured interview questions for the medication taker interviews (study 2).[PDF File (Adobe PDF File), 90 KB - humanfactors_v10i1e40173_app2.pdf]

Multimedia Appendix 3
Semistructured interview questions for the health care provider interviews (study 3).[PDF File (Adobe PDF File), 70 KB - humanfactors_v10i1e40173_app3.pdf]

References


Abbreviations
- AI: artificial intelligence
- FDA: Food and Drug Administration
- HIPAA: Health Insurance Portability and Accountability Act of 1996
- IoT: Internet of Things
- ML: machine learning
- PDR: Pre-Departure Reminder
- PN: Proximity Notification
- PN-Care: Proximity Notifications to Caregiver
- PrEP: pre-exposure prophylaxis
- RCS: Routine Change Detection
- RCS-Team: Routine Change Detection Supported with Healthcare Providers
- WHO: World Health Organization

©Jacob T Biehl, Ravi Patel, Adam J Lee. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 04.07.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information,
a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Workload, Usability, and Engagement with a Mobile App Supporting Video Observation of Methadone Take-Home Dosing: Usability Study

Bulat Idrisov¹, MSc, MD; Kevin A Hallgren², PhD; Alyssa Michaels³, BA, MSc; Sean Soth⁴, BA; James Darnton⁴,⁵, MD; Paul Grekin²,⁴, MD; Steve Woolworth¹, PhD; Andrew J Saxon²,⁶, MD; Judith I Tsui⁵, MPH, MD

¹Department of Health Systems and Population Health, University of Washington, Seattle, WA, United States
²Department of Psychiatry and Behavioral Sciences, University of Washington School of Medicine, Seattle, WA, United States
³Division of HIV, Infectious Diseases, and Global Medicine, Department of Medicine, University of California San Francisco, San Francisco, CA, United States
⁴Evergreen Treatment Services, Seattle, WA, United States
⁵Division of General Internal Medicine, University of Washington, Seattle, WA, United States
⁶Center of Excellence in Substance Addiction Treatment and Education, Veterans Affairs Puget Sound Health Care System, Seattle, WA, United States

Corresponding Author:
Bulat Idrisov, MSc, MD
Department of Health Systems and Population Health
University of Washington
3980 15th Ave NE
Seattle, WA, 98195-1621
United States
Phone: 1 2067937662
Email: idrisov@uw.edu

Abstract

Background: Methadone, a cornerstone of opioid use disorder treatments for many decades, is an essential tool for combatting the opioid epidemic. However, requirements for observing methadone dosing in person through direct observed therapy (DOT) impose significant barriers for many patients. Digital technology can facilitate remote DOT, which could reduce barriers to methadone treatment. Currently, there are limited data on the usability of such technology among patients and counselors in methadone treatment settings.

Objective: The primary objective of this study was to assess the workload, usability, and engagement of a video-based DOT mobile app for patients with opioid use disorder receiving methadone treatment. The secondary objective was to assess the workload, usability, and engagement of the provider-facing app portal used by counselors.

Methods: Patients (n=12) and counselors (n=3) who previously tried video DOT for methadone through a smartphone app in an opioid treatment program participated in usability testing sessions. Participants completed essential tasks for video DOT, then provided ratings of workload (NASA Task Load Index), usability (modified System Usability Scale), and engagement (modified Engagement Scale) with the core features of the video DOT program.

Results: Patients and counselors reported low mental, physical, and temporal demands, successful performance, low effort, and low frustration associated with activities. Patients reported high usability (mean 85, SD 9.5) and engagement (mean 3.8, SD 1.1); counselors reported moderate usability (mean 43.3, SD 17.7) and engagement (mean 2.81, SD 0.63).

Conclusions: A mobile health app that facilitates video-based DOT for methadone required a low workload for patients and counselors and was highly usable for patients in an opioid treatment program; however, there are opportunities to improve usability and engagement for the counselor-facing portal.

(JMIR Hum Factors 2023;10:e42654) doi:10.2196/42654

KEYWORDS
addiction; direct observed therapy; health app; methadone; mHealth; mobile app; mobile health; opioid; smartphone app; substance use; usability; user engagement; user testing; workload
Introduction

Opioid use disorder (OUD) remains a major cause of mortality in the United States [1]. Methadone is 1 of 3 OUD pharmacotherapies approved by the US Food and Drug Administration but requires frequent in-person observed dosing (ie, direct observed therapy [DOT]) at a federally certified opioid treatment program (OTP) to mitigate the risks of medication diversion and overdose. The requirement of DOT can impose barriers for patients and limit access to treatment [2].

Mobile health (mHealth) technology has the potential to help reduce barriers to methadone treatment [3]. For example, smartphones allow patients to video-record themselves taking methadone at home or send messages to clinical providers, which can reduce the need for frequent visits to an OTP for DOT. However, in-person DOT remains the standard by regulation in OTPs [4].

In response to concerns about respiratory illness transmission during the COVID-19 pandemic, a large OTP agency with 3 separate sites in Washington state conducted a pilot program between April and August 2020 aimed at reducing the need for in-person DOT and doing remote screenings of COVID-19 symptoms. During the pilot, a subset of patients was invited to use the Emocha mHealth app to facilitate video-based DOT for all methadone take-home doses along with COVID-19 symptom screening completed with each video DOT submission.

In a study (Hallgren et al [5]) describing the clinical pilot, we showed that patient adherence to video DOT varied, but on average, video DOT significantly increased the number of days of observed methadone dosing and most patients received increased methadone take-home dosing privileges due to their ability to demonstrate treatment stability. However, the direct usability of the video DOT app has yet to be tested with patients or counselors in methadone treatment settings.

For this study, patients and counselors who participated in the pilot program were invited to participate in an evaluation of the app’s overall usability. We hypothesized that the mobile app would have favorable workload, usability, and engagement for patients with OUD receiving methadone treatment, including for patients who had higher versus lower adherence to the app during the original pilot program. We also hypothesized that counselors would report favorable workload, usability, and engagement for the provider-facing app portal.

Methods

Study Design and Sample

Patients and counselors who participated in the original pilot program and were still receiving care or employed by the methadone treatment program were invited to participate in the usability study between May and August 2021. Patients and counselors were invited using phone calls, letters, and flyers distributed at the OTP. Recruited counselors were also encouraged to refer participants from the pilot to take part in this usability study. Efforts were made to recruit patients who in the original clinical pilot had low adherence (less than 18 video uploads), medium adherence (18-45 video uploads), and high adherence (more than 45 video uploads), defined by terciles of video uploads. Additional information describing the original pilot program and outcomes was reported by Hallgren et al [5].

Testing Procedures

Participants completed a single usability session conducted 1-on-1 with a research coordinator in a private setting following a standardized protocol. After providing informed consent, participants completed a demographic questionnaire. The research assistant (RA) administered usability testing tasks and questionnaires and recorded data into REDCap. To complete usability testing tasks, patients either used the mobile app on a study phone or downloaded the app to their phone.

In the first session of the usability study, our RA provided verbal instructions and prompts as participants engaged in each of the tasks. This approach was chosen because, unlike a self-help app where users typically interact with the app without guidance, the Emocha app is designed for users who receive instructions from health care staff on how to complete specific tasks. Therefore, providing instructions during the usability testing accurately reflects the intended user experience and was seen as the most appropriate methodology for this study. One illustrative example of the scripts used by our RAs during the study is as follows: “Please open the mobile app and log on using the provided username and password.”

Patient participants were asked to complete 5 tasks that were determined by the study team to be the most important for successful video DOT: logging into the account, completing a COVID-19 symptom screener, uploading a video of themselves simulating methadone ingestion, sending and checking messages to a counselor, and accessing and reviewing a calendar showing methadone adherence. Counselors completed usability testing tasks on the provider-facing web portal using a study computer if the visit was in person or through their own work or personal computer if the visit was conducted remotely through Zoom. Counselor participants were asked to complete 5 tasks, that is, add a new patient, review 1 patient video, change a patient’s video regimen time and number of uploads, send and check messages, and check the patient “adherence calendar.”

The research coordinator timed each task and observed whether it was completed successfully. Participants provided NASA Task Load Index (NASA-TLX) ratings after each activity. After all activities were completed, participants completed the System Usability Scale (SUS) and User Engagement Scale-Short Form (UES-SF), described below.

Ethics Approval

All procedures were approved by the University of Washington Institutional Review Board (review number STUDY00011142).

Measures

NASA-TLX Measure

The NASA-TLX is a validated measure [6] of the cognitive workload required to complete a task. Participants self-report the mental demand, physical demand, temporal demand, performance, effort, and frustration associated with each of the 5 activities completed during the usability testing session on a...
visual analog scale of 0-100. An overall task load index was computed as the unweighted mean rating across all 5 activities [7]. We derived the following cutoffs to interpret mean workload: <33 for low workload, 33-66 for moderate workload, and >66 for high workload. These cutoffs were informed by Patel et al [8], who reviewed workload ratings for 21 electronic medication adherence apps and found an average workload of 50 (SD 26) with some of the least workload-heavy products having mean ratings of around 29.

**SUS Measure**

The SUS is a validated self-report usability measure [9]. It has 10 statements (5 positively framed and 5 negatively framed) that are rated on a 5-point scale completed at the end of the testing session. Total scores were calculated following standard instructions [10] to produce a total score ranging from 0 to 100, with higher scores indicating greater usability. Similar to our NASA-TLX scores, we interpreted mean scores <33 as low usability, 33-77 as moderate usability, and >77 as high usability. These cutoffs were informed by previous studies showing that the most usable medication adherence products had mean SUS scores of about 78 (SD 15) and the least usable medication adherence products had mean scores of around 28 (SD 21) [8].

**UES-SF Measure**

User engagement reflects the depth of cognitive, temporal, affective, and behavioral investment when interacting with a digital system [11] and was measured using the UES-SF [12]. The UES-SF is a 12-item self-report measure with 4 dimensions reflecting focused attention, perceived usability, aesthetic appeal, and reward factor; the latter subscale combines a felt sense of novelty, involvement, and endurability experienced while interacting with the digital system. Each question is answered on a 5-point rating scale. Following recommendations by O’Brien [11] and O’Brien et al [12], we calculated mean scores for each subscale and an overall engagement score reflecting the mean rating of all 12 items (negative engagement items were reverse coded). For this analysis, an average score higher than 3.5 would indicate high engagement.

**Analytic Approach**

Descriptive statistics characterized the patient sample. Workload, usability, and engagement measures were analyzed descriptively by computing means and 95% CIs of composite indices within the patient and counselor cohorts. Additional descriptive analyses were performed within patient subgroups who, during the original clinical pilot, had video DOT adherence that was considered low to moderate (n=8 patients) and high (n=4 patients) during the first 60 days from enrollment.

**Results**

**Description of Sample**

The study recruited and enrolled 12 of the 60 patients who participated in the clinical pilot (2=low adherence, 6=medium adherence, and 4=high adherence) during the first 60 days from enrollment. Table 1 describes the patient participants in the usability study. Demographics for counselors are not reported due to the small number of those participants. On average, patients were in their late 40s and most were male and White.
Table 1. Description of study patients who participated in the study. Demographics for counselors are not reported to preserve confidentiality, given the small sample size (n=3).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Patients (n=12), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;30</td>
<td>0 (0)</td>
</tr>
<tr>
<td>30-49</td>
<td>5 (42)</td>
</tr>
<tr>
<td>50-64</td>
<td>7 (58)</td>
</tr>
<tr>
<td>≥65</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>9 (75)</td>
</tr>
<tr>
<td>Female</td>
<td>3 (25)</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
</tr>
<tr>
<td>American Indian or Alaskan Native</td>
<td>2 (17)</td>
</tr>
<tr>
<td>Asian or Asian American</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Black or African American</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Native Hawaiian or Pacific Islander</td>
<td>1 (8)</td>
</tr>
<tr>
<td>White</td>
<td>9 (75)</td>
</tr>
<tr>
<td>Unknown or another race</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Not Hispanic or Latino</td>
<td>12 (100)</td>
</tr>
<tr>
<td>Unknown</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Homeless</strong></td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

*Nonexclusive category.

Task Completion

All 5 activities were successfully completed by all participants, with the exception that 1 patient did not successfully log into the app. It took an average 1.6 minutes for patients to complete each of the 5 activities. The most time-consuming activity was logging into the account, which took a mean of 3 minutes; however, this mean was greatly affected by 2 outlier participants who took 9.4 and 13.8 minutes to complete the task. Of the 10 remaining participants, 8 completed the task in less than 2 minutes, and 2 completed the task in 2-2.5 minutes. The RA observed that some participants took longer to complete the login task because of problems not directly related to the software. For example, 1 patient participant engaged in conversation while attempting to log in, which prolonged the process. Another participant entered an incorrect test password, resulting in failed login attempts. In another instance, a slow-performing phone impacted and slowed the process and appeared to create login failures. The second most time-demanding activity was sending and checking messages (1.9 minutes on average). For clinicians, it took an average 1.4 minutes to complete each of the 5 activities. The most time-demanding activities were changing a patient’s video regimen time and number of uploads and reviewing 1 patient video (both 2.2 minutes on average).

Workload (NASA-TLX)

Overall, patients and counselors reported low mental, physical, and temporal demands, successful performance, low effort, and low frustration associated with activities. However, counselors reported somewhat higher demands across all categories and activities. Results for the 2 video DOT adherence subgroups and counselors are presented in Tables 2 and 3.
## Activities (NASA Task Load Index) for patients (n=12).

<table>
<thead>
<tr>
<th>Task</th>
<th>Description</th>
<th>Task 1: log into account</th>
<th>Task 2: symptom screening completion</th>
<th>Task 3: upload a video mimicking ingestion of methadone</th>
<th>Task 4: send and check a message</th>
<th>Task 5: check progress in the adherence calendar, and state percentage of videos uploaded</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of people who completed the task successfully*, n (%)</td>
<td>11 (92)</td>
<td>12 (100)</td>
<td>12 (100)</td>
<td>12 (100)</td>
<td>12 (100)</td>
</tr>
<tr>
<td></td>
<td>Time to complete task (seconds), mean (SD)</td>
<td>181 (251)</td>
<td>50 (38)</td>
<td>80.7 (65)</td>
<td>115 (88)</td>
<td>62.8 (71)</td>
</tr>
<tr>
<td></td>
<td>Mental demand (from 0 to 100); “How mentally demanding was the activity?” mean (SD)</td>
<td>6 (11)</td>
<td>5 (11)</td>
<td>10 (15)</td>
<td>6.5 (7)</td>
<td>9.2 (16)</td>
</tr>
<tr>
<td></td>
<td>Physical demand (from 0 to 100); “How physically demanding was the activity?” mean (SD)</td>
<td>3.6 (8)</td>
<td>3 (7)</td>
<td>6.5 (6)</td>
<td>2.7 (3)</td>
<td>5.9 (17)</td>
</tr>
<tr>
<td></td>
<td>Temporal demand (from 0 to 100); “How hurried or rushed was the pace of the activity?” mean (SD)</td>
<td>7.6 (15)</td>
<td>9.7 (19)</td>
<td>8.2 (15)</td>
<td>5.9 (14)</td>
<td>9.6 (19)</td>
</tr>
<tr>
<td></td>
<td>Performance (from 0 to 100); “How successful were you in accomplishing what you were asked to do?” mean (SD)</td>
<td>81.2 (37)</td>
<td>97.5 (7)</td>
<td>97.5 (6)</td>
<td>90.3 (28)</td>
<td>94 (17)</td>
</tr>
<tr>
<td></td>
<td>Effort (from 0 to 100); “How hard did you have to work to accomplish your level of performance?” mean (SD)</td>
<td>6 (14)</td>
<td>7.5 (20)</td>
<td>13.3 (25)</td>
<td>4.7 (5)</td>
<td>8.8 (23)</td>
</tr>
<tr>
<td></td>
<td>Frustration (from 0 to 100); “How insecure, discouraged, irritated, stressed, and annoyed were you?” mean (SD)</td>
<td>4 (8)</td>
<td>4.75 (14)</td>
<td>6.7 (8)</td>
<td>4.5 (7)</td>
<td>1.6 (3)</td>
</tr>
</tbody>
</table>

*Task 1 successful completion: logging into the mobile app using the username and password provided. Task 2 successful completion: indicating both cough and fever on the symptom screener. Task 3 successful completion: recording video with all instructions followed and submitting video. Task 4 successful completion: checking the message and replying to the question in the mobile app chat function. Task 5 successful completion: locating the adherence calendar and stating the percentage.
Table 3. Activities (NASA Task Load Index) for counselors (n=3).

<table>
<thead>
<tr>
<th>Task</th>
<th>Number of people completed task successfully, n (%)</th>
<th>Time to complete task (seconds), median (min, max)</th>
<th>Mental demand (from 0 to 100): “How mentally demanding was the activity?” median (min, max)</th>
<th>Physical demand (from 0 to 100): “How physically demanding was the activity?” median (min, max)</th>
<th>Temporal demand (from 0 to 100): “How hurried or rushed was the pace of the activity?” median (min, max)</th>
<th>Performance (from 0 to 100): “How successful were you in accomplishing what you were asked to do?” median (min, max)</th>
<th>Effort (from 0 to 100): “How hard did you have to work to accomplish your level of performance?” median (min, max)</th>
<th>Frustration (from 0 to 100): “How insecure, discouraged, irritated, stressed, and annoyed were you?” median (min, max)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1: adding a new patient</td>
<td>2 (67)</td>
<td>99 (73, 144)</td>
<td>25 (10, 60)</td>
<td>25 (1, 50)</td>
<td>10 (1, 50)</td>
<td>25 (0, 100)</td>
<td>25 (10, 70)</td>
<td>26 (15, 60)</td>
</tr>
<tr>
<td>Task 2: reviewing 1 patient video</td>
<td>3 (100)</td>
<td>128 (53, 213)</td>
<td>60 (10, 75)</td>
<td>10 (1, 50)</td>
<td>10 (1, 40)</td>
<td>50 (40, 75)</td>
<td>25 (25, 60)</td>
<td>60 (25, 75)</td>
</tr>
<tr>
<td>Task 3: changing a patient’s regimen time and number of uploads</td>
<td>1 (33)</td>
<td>95 (59, 247)</td>
<td>50 (30, 70)</td>
<td>30 (1, 50)</td>
<td>30 (1, 50)</td>
<td>50 (0, 90)</td>
<td>60 (30, 75)</td>
<td>60 (50, 75)</td>
</tr>
<tr>
<td>Task 4: sending and checking a message</td>
<td>2 (67)</td>
<td>91 (27, 108)</td>
<td>15 (5, 60)</td>
<td>5 (1, 50)</td>
<td>1 (0, 50)</td>
<td>100 (0, 100)</td>
<td>15 (5, 75)</td>
<td>10 (5, 60)</td>
</tr>
<tr>
<td>Task 5: checking the patient “adherence calendar” and stating percentage of adherent video</td>
<td>1 (33)</td>
<td>71 (68, 77)</td>
<td>69 (5, 70)</td>
<td>5 (0, 50)</td>
<td>5 (1, 50)</td>
<td>0 (0, 100)</td>
<td>50 (20, 50)</td>
<td>60 (10, 90)</td>
</tr>
</tbody>
</table>

Usability (SUS)

Mean SUS scores reflected high usability for patients (mean 85, SD 9.5). Usability was also high for the patient subgroups with low to moderate video DOT adherence (mean 87.5, SD 7.9) and high adherence (mean 79.3, SD 11.2; Table 4), a nominal difference that was not statistically significant (mean difference 8.5, 95% CI –3.8 to 20.8). In contrast, usability ratings for counselors were considerably lower (mean 43.3, SD 17.7).
Table 4. Usability score for patients overall (n=12) and by adherence to the app group and counselors (n=3).

<table>
<thead>
<tr>
<th>Questions</th>
<th>Patients (n=12), mean (SD)</th>
<th>Patients with low to moderate adherence to video DOT&lt;sup&gt;a&lt;/sup&gt; during clinical pilot (n=8), mean (SD)</th>
<th>Patients with high adherence to video DOT during clinical pilot (n=4), mean (SD)</th>
<th>Counselors (n=3), mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>“I think I would like to use this mobile app (web portal) frequently along with my methadone treatment” (scored from 1 to 5)</td>
<td>4.3 (1)</td>
<td>4.6 (0.7)</td>
<td>3.7 (1.5)</td>
<td>2.3 (0.5)</td>
</tr>
<tr>
<td>“I found the mobile app (web portal) unnecessarily complex” (scored from 1 to 5)</td>
<td>1.6 (0.5)</td>
<td>1.3 (0.5)</td>
<td>2.0 (0)</td>
<td>2.6 (1.1)</td>
</tr>
<tr>
<td>“I thought the mobile app (web portal) was easy to use” (scored from 1 to 5)</td>
<td>4.3 (1.2)</td>
<td>4.2 (1.4)</td>
<td>4.2 (0.5)</td>
<td>3.1 (1)</td>
</tr>
<tr>
<td>“I think that I would need the support of a technical person to be able to use this mobile app (web portal)” (scored from 1 to 5)</td>
<td>1.9 (1)</td>
<td>2.1 (1.3)</td>
<td>1.7 (0.5)</td>
<td>4.3 (0.6)</td>
</tr>
<tr>
<td>“I found the various functions in this mobile app (web portal) were well integrated” (scored from 1 to 5)</td>
<td>4.5 (0.7)</td>
<td>4.7 (0.5)</td>
<td>4.0 (0.8)</td>
<td>3.3 (1.1)</td>
</tr>
<tr>
<td>“I thought there was too much inconsistency in this mobile app (web portal)” (scored from 1 to 5)</td>
<td>1.5 (0.5)</td>
<td>1.4 (0.5)</td>
<td>1.7 (0.5)</td>
<td>2.3 (0.6)</td>
</tr>
<tr>
<td>“I would imagine that most people would learn to use this mobile app (web portal) very quickly” (scored from 1 to 5)</td>
<td>4.7 (0.5)</td>
<td>4.7 (0.4)</td>
<td>4.5 (0.5)</td>
<td>2.1 (1)</td>
</tr>
<tr>
<td>“I found the mobile app (web portal) very awkward to use” (scored from 1 to 5)</td>
<td>1.5 (0.5)</td>
<td>1.5 (0.5)</td>
<td>1.5 (0.6)</td>
<td>2.6 (1.1)</td>
</tr>
<tr>
<td>“I felt very confident using the mobile app (web portal)” (scored from 1 to 5)</td>
<td>4.6 (0.7)</td>
<td>4.7 (0.5)</td>
<td>4.2 (0.9)</td>
<td>2.6 (1.1)</td>
</tr>
<tr>
<td>“I needed to learn a lot of things before I could get going with this mobile app (web portal)” (scored from 1 to 5)</td>
<td>3.9 (0.8)</td>
<td>1.8 (1)</td>
<td>2.0 (0)</td>
<td>4.1 (1)</td>
</tr>
<tr>
<td>Overall SUS&lt;sup&gt;b&lt;/sup&gt; score on a 0 to 100 normalized scale</td>
<td>85 (9.5)</td>
<td>87.5 (7.9)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>79.3 (11.2)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>43.3 (17.7)</td>
</tr>
</tbody>
</table>

<sup>a</sup>DOT: direct observed therapy.

<sup>b</sup>SUS: System Usability Scale.

<sup>c</sup>The mean difference between the 2 groups of users was 8.5 (95% CI –3.804 to 20.804).

**User engagement (UES-SF)**

User engagement was high for patients (mean 3.8, SD 1.1). User engagement was high for the low to moderate video DOT adherence subgroup (mean 3.9, SD 1.2), but it was lower for the high adherence subgroup (mean 2.8, SD 1.1), a nominal difference that was not statistically significant (mean difference 1.1, 95% CI –0.5 to 2.7). Results for specific domain categories are described in Table 5. For counselors, engagement was lower (mean 2.8, SD 0.6), particularly for the reward and perceived usability domains.
Table 5. Four domains of the User Engagement Scale–Short Form for patients overall (n=12), by adherence to the app group, and for counselors (n=3).

<table>
<thead>
<tr>
<th>Questions and subsequent domain scored from 1 to 5</th>
<th>Patients (n=12), mean (SD)</th>
<th>Patients with low to moderate adherence to video DOT&lt;sup&gt;a&lt;/sup&gt; during clinical pilot (n=8), mean (SD)</th>
<th>Patients with high adherence to video DOT during clinical pilot (n=4), mean (SD)</th>
<th>Difference between 2 patient groups of users, mean (95% CI)</th>
<th>Counselors (n=3), mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. “I lost myself in this experience”</td>
<td>2 (0.7)</td>
<td>2 (0.8)</td>
<td>4 (0.8)</td>
<td>N/A&lt;sup&gt;b&lt;/sup&gt;</td>
<td>3.7 (0.58)</td>
</tr>
<tr>
<td>2. “The time I spent using mobile app (web portal) just slipped away”</td>
<td>2.8 (1.1)</td>
<td>2.9 (1.2)</td>
<td>3.3 (1)</td>
<td>N/A</td>
<td>4.0 (0)</td>
</tr>
<tr>
<td>3. “I was absorbed in this experience”</td>
<td>3.8 (1.3)</td>
<td>3.6 (1.5)</td>
<td>2.0 (0.8)</td>
<td>N/A</td>
<td>2.3 (0.58)</td>
</tr>
<tr>
<td>Mean of items 1-3, measuring the “focused attention” domain. Items scored as the following: strongly disagree=1, disagree=2, neither disagree nor agree=3, agree=4, and strongly agree=5.</td>
<td>2.9 (0.7)</td>
<td>2.8 (0.9)</td>
<td>3.1 (0.3)</td>
<td>0.3 (–0.752 to 1.352)</td>
<td>3.3 (0)</td>
</tr>
<tr>
<td>4. “I felt frustrated while using this mobile app (web portal)”</td>
<td>4.4 (0.67)</td>
<td>4.6 (0.5)</td>
<td>4.0 (0.8)</td>
<td>N/A</td>
<td>2.3 (0.58)</td>
</tr>
<tr>
<td>5. “I found this mobile app (web portal) confusing to use”</td>
<td>4.2 (1.14)</td>
<td>4.3 (1.4)</td>
<td>4.3 (0.5)</td>
<td>N/A</td>
<td>3.1 (0)</td>
</tr>
<tr>
<td>6. “Using this mobile app (web portal) was taxing”</td>
<td>4.2 (0.87)</td>
<td>4.6 (0.5)</td>
<td>3.5 (1)</td>
<td>N/A</td>
<td>2.0 (0)</td>
</tr>
<tr>
<td>Mean of items 4-6, measuring the “perceived usability” domain. Scored as the following: strongly disagree=5, disagree=4, neither disagree nor agree=3, agree=2, and strongly agree=1.</td>
<td>4.3 (0.63)</td>
<td>4.5 (0.5)</td>
<td>3.9 (0.7)</td>
<td>0.6 (–0.174 to 1.374)</td>
<td>2.4 (0.51)</td>
</tr>
<tr>
<td>7. “This mobile app was (web portal) attractive”</td>
<td>3.5 (0.67)</td>
<td>3.5 (0.8)</td>
<td>2.5 (0.6)</td>
<td>N/A</td>
<td>3.1 (0)</td>
</tr>
<tr>
<td>8. “This mobile app (web portal) was aesthetically appealing”</td>
<td>3.7 (0.78)</td>
<td>3.4 (0.7)</td>
<td>1.8 (0.5)</td>
<td>N/A</td>
<td>2.6 (1.53)</td>
</tr>
<tr>
<td>9. “This mobile app (web portal) appealed to my visual senses”</td>
<td>3.3 (0.89)</td>
<td>3.3 (0.9)</td>
<td>2.5 (1)</td>
<td>N/A</td>
<td>3.3 (1.53)</td>
</tr>
<tr>
<td>Mean of items 7-9, measuring the “aesthetic appeal” domain. Scored as the following: strongly disagree=1, disagree=2, neither disagree nor agree=3, agree=4, and strongly agree=5.</td>
<td>3.5 (0.7)</td>
<td>3.4 (0.8)</td>
<td>2.3 (0.6)</td>
<td>1.1 (0.083 to 2.117)</td>
<td>3.1 (1.33)</td>
</tr>
<tr>
<td>10. “Using this mobile app (web portal) was worthwhile”</td>
<td>4.6 (0.51)</td>
<td>4.8 (0.5)</td>
<td>1.8 (0.5)</td>
<td>N/A</td>
<td>2.7 (1.15)</td>
</tr>
<tr>
<td>11. “My experience was rewarding”</td>
<td>4.7 (0.65)</td>
<td>4.9 (0.4)</td>
<td>1.8 (1)</td>
<td>N/A</td>
<td>2.7 (0.58)</td>
</tr>
<tr>
<td>12. “I felt interested in this experience”</td>
<td>4.2 (0.62)</td>
<td>4.5 (0.5)</td>
<td>2.3 (0.5)</td>
<td>N/A</td>
<td>2.0 (0)</td>
</tr>
<tr>
<td>Mean of items 10-12, measuring the “reward” domain. Scored as the following: strongly disagree=1, disagree=2, neither disagree nor agree=3, agree=4, and strongly agree=5.</td>
<td>4.5 (0.5)</td>
<td>4.7 (0.4)</td>
<td>1.9 (0.6)</td>
<td>2.8 (2.16 to 3.44)</td>
<td>2.4 (0.51)</td>
</tr>
<tr>
<td>Total score for user engagement (an average score higher than 3.5 indicates high engagement)</td>
<td>3.8 (1.12)</td>
<td>3.9 (1.2)</td>
<td>2.8 (1.1)</td>
<td>1.1 (–0.498 to 2.698)</td>
<td>2.8 (0.63)</td>
</tr>
</tbody>
</table>

<sup>a</sup>DOT: direct observed therapy.

<sup>b</sup>N/A: not applicable.
Discussion

Overview
Methadone is a life-saving medication for patients with OUD, but requirements for in-person DOT can impose significant barriers to treatment. This study found that an mHealth platform that facilitates video-based DOT and COVID-19 symptom screening through smartphones required low workload and had high usability and engagement for patients with OUD receiving methadone and required low workload and had moderate usability and engagement for methadone treatment counselors. However, counselors scored lower across all instruments. Results indicated that the 5 most critical functions of the app could almost always be completed by patients and counselors and that these tasks were associated with low cognitive workload, high usability, and high user engagement, including for patients with low to moderate adherence and high adherence in the original pilot.

Results suggest that video DOT can be usable for patients with OUD in methadone treatment. The strong performance observed for the study cohort, including in patients with low to moderate adherence in the original pilot, suggests that usability was unlikely to be a significant barrier to adherence with video DOT and that other barriers may have contributed more to variability in video DOT adherence. Contextual factors influencing experiences using the app are currently being explored in a separate qualitative study.

Further investigation is necessary to determine the reasons behind counselors’ lower ratings of the mHealth platform’s usability and engagement compared to patients. It is worth noting that the counselor-facing portal had different features than the patient-facing app. For example, the tasks of counselors were different from patients, as patients submitted videos that counselors then reviewed and approved. We speculate that counselor usability ratings could have been impacted by software and clinical workflow issues. These factors need further investigation, especially given the small sample size in our study. However, possible solutions to address these issues could include providing additional training and support to counselors to enable them to feel ownership, ease, and mastery of their role, such as adding more advanced features or customization options to the platform and conducting more rigorous testing and evaluation of the counselor-facing portal to identify and address any specific usability issues. Future work could also examine how to improve the integration of video DOT with the counselors’ clinic routines and existing workloads, including by identifying ways to minimize the potential impacts of such systems on their clinical routines.

Our study has limitations, such as the small sample size, especially for counselors. With 2 of 5 counselors missing, the findings might be unrepresentative and biased. Although descriptive analyses can be conducted with small sample sizes, the precision of results is limited, and we could not perform subgroup analyses to evaluate usability within important patient subgroups (eg, patients experiencing homelessness or higher-severity OUD). There may have been sampling bias, as we were only able to recruit patients and counselors who were in the original pilot program and were still in the clinic over 1 year later. We also recruited only 2 patients with low adherence in the original clinical pilot, which may introduce a bias toward more favorable results, as one would hypothesize that patients with low adherence might be more likely to experience problems with usability (however, several usability ratings were nominally higher for participants with low to medium adherence compared to patients with high adherence). Our study also has strengths, including its focus on analyzing a novel method for DOT for methadone and usability testing with people who may often be overlooked in technology development efforts.

Conclusion
Little knowledge exists on the usability of mHealth apps for patients and counselors in methadone treatment. This study narrows the knowledge gap by providing information on workload, usability, and engagement with an mHealth app delivered through smartphones for video observation of methadone home dosing and COVID-19 symptom screening. The study demonstrated that a mHealth app to facilitate video-based DOT of methadone was unlikely to create a heavy workload for patients and counselors. Furthermore, there were no trends to suggest that adherence to the app in the original clinical pilot was related to workload, usability, or engagement, indicating that factors unrelated to usability may have impacted adherence when the app was used during the clinical pilot. Although the app was well received by patients, the study highlights opportunities to improve and further investigate usability for counselors, perhaps by improving training or care integration.

Acknowledgments
Research reported in this publication was supported by the National Institute on Drug Abuse (NIDA) (award R41DA053081) and the National Institute on Alcohol Abuse and Alcoholism (NIAAA) (award K01AA024796). The content is solely the responsibility of the authors and does not necessarily represent the official views of NIDA, NIAAA, or the National Institutes of Health. AJS has received travel support from Alkermes Inc, payment for serving on an advisory board from Indivior Inc, and royalties from UpToDate Inc. We would like to thank Kendra Blalock and Zachery Schramm for their contributions to study start-up and selection of usability measurements, respectively, and Michelle Peavy for assistance in securing grant funding. We would like to thank Lauren Brown, Sebastian Seiguer, and other staff at Emocha. The company had no role in data collection, interpretation of data, and writing of the manuscript.
Conflicts of Interest
None declared.

References

Abbreviations
DOT: direct observed therapy
FDA: Food and Drug Administration
mHealth: mobile health
NASA-TLX: NASA Task Load Index
OTP: opioid treatment program
OUD: opioid use disorder
RA: research assistant
SUS: System Usability Scale
UES-SF: User Engagement Scale–Short Form

Edited by A Kushniruk; submitted 26.09.22; peer-reviewed by T Wray, J Pytell; comments to author 26.02.23; revised version received 02.05.23; accepted 25.05.23; published 13.07.23.

Please cite as:
Workload, Usability, and Engagement with a Mobile App Supporting Video Observation of Methadone Take-Home Dosing: Usability Study.
JMIR Hum Factors 2023;10:e42654
URL: https://humanfactors.jmir.org/2023/10/e42654
doi:10.2196/42654
PMID:37440298
Digitally Based Blood Pressure Self-Monitoring Program That Promotes Hypertension Self-Management and Health Education Among Patients With Low-Income: Usability Study

Jacqueline Yareli Poblete¹, MPH; Natalie Lauren Vawter¹, MPH; Sydney Virginia Lewis¹; Earl Marc Felisme¹; Paloma Adriana Mohn¹; Jennifer Shea¹, MPH; Adam William Northrup¹, MS; Jie Liu¹, MPH, MD; Tala Al-Rousan², MPH, MD; Job Gideon Godino¹,², PhD

¹Laura Rodriguez Research Institute, Family Health Centers of San Diego, San Diego, CA, United States
²Herbert Wertheim School of Public Health and Human Longevity Science, University of California San Diego, San Diego, La Jolla, CA, United States

**Corresponding Author:**
Job Gideon Godino, PhD
Laura Rodriguez Research Institute
Family Health Centers of San Diego
LRRI, 3rd Fl
1750 Fifth Ave
San Diego, CA, 92101
United States
Phone: 1 619 515 2344
Email: jobjg@fhcsd.org

**Abstract**

**Background:** According to evidence-based clinical guidelines, adults with hypertension are advised to self-monitor their blood pressure (BP) twice daily. Self-measured BP monitoring is a recommended strategy for improving hypertension management.

**Objective:** We aimed to determine the feasibility and acceptability of a digitally based BP self-monitoring program that promotes hypertension self-management and health education among low-income patients. We hypothesized that the program would be highly feasible and acceptable and that at least 50% of the patients would use the monitor at the rate required for the reimbursement of the device’s cost (16 days of measurements in any 30-day period).

**Methods:** Withings BPM Connect was deployed to patients at Family Health Centers of San Diego. Program elements included training, SMS text message reminders, and physician communication. Compliance, use, mean BP, and BP control status were calculated. A Kaplan-Meier time-to-event analysis was conducted to compare time to compliance between a strict definition (≥16 days in any rolling 30-day window) and a lenient definition (≥1 day per week for 4 consecutive weeks). A log-rank test was performed to determine whether the difference in time to compliance between the definitions was statistically significant. Mean systolic BP (SBP) and diastolic BP (DBP) before the intervention and after the intervention and mean change in SBP and DBP across patients were calculated. Paired sample t tests (2-tailed) were performed to assess the changes in SBP and DBP from before to after the intervention.

**Results:** A total of 179 patients received the monitors. The mean changes in SBP and DBP from before to after the intervention were +2.62 (SE 1.26) mm Hg and +3.31 (SE 0.71) mm Hg, respectively. There was a statistically significant increase in both SBP and DBP after the intervention compared with before the intervention (P=.04 and P<.001). At the first and last measurements, 37.5% (63/168) and 48.8% (82/168) of the patients had controlled BP, respectively. During the observation period, 83.3% (140/168) of the patients had at least 1 controlled BP measurement. Use decreased over time, with 53.6% (90/168) of the patients using their monitor at week 2 and only 25% (42/168) at week 11. Although only 25.6% (43/168) achieved the strict definition of compliance, 42.3% (71/168) achieved the lenient definition of compliance. The median time to compliance was 130 days for the strict definition and 95 days for the lenient definition. The log-rank test showed a statistically significant difference in time to compliance between the compliance definitions (P<.001). Only 26.8% (45/168) complied with the measurement rate that would result in device cost reimbursement.

**Conclusions:** Few patients used the monitors at a rate that would result in reimbursement, raising financial feasibility concerns. Plans for sustaining costs among low-income patients need to be further evaluated.
Introduction

Background

High blood pressure (BP), or hypertension, is one of the most prevalent health issues in the United States, affecting almost half of the adults (47% or 116 million) [1]. In the United States, 1 in 5 adults is unaware that they have hypertension, likely owing to a lack of symptoms [2], and only one-quarter of US adults with hypertension have the condition under control (<140/90 mm Hg) [1]. Untreated and uncontrolled hypertension raises the risk of heart disease, stroke, kidney damage, and other complications [3]. In 2020, hypertension was a main or contributing factor in >670,000 deaths in the United States [4]. Along with its impacts on morbidity and mortality, hypertension has a substantial economic burden, costing the United States between US $131 and US $198 billion per year (including the cost of health care services and BP medications) [5]. Patients with hypertension also have increased costs (>US $2000 more per year) compared with those without hypertension, including nearly double the annual prescription medication costs and 2.5 times the annual hospital expenses [5].

According to the Health Resources and Services Administration, 30 million Americans (1 in every 3 living in poverty and 1 in every 7 of a racial or ethnic minority group) receive medical care from federally qualified health centers (FQHCs) [6]. Although FQHCs provide preventive chronic disease and primary care services to medically underserved populations, uninsured individuals who visit an FQHC can still be charged for their care. The lack of health insurance is one of the most significant barriers to effective hypertension management among US adults [7]. Therefore, low-income and uninsured populations are the most susceptible to inadequate hypertension treatment and management. The current limited information on hypertension management in FQHCs highlights a need for further research in this patient population.

Hypertension management, including medication and lifestyle changes, can significantly reduce morbidity and mortality rates. Self-measured BP (SMBP) monitoring with clinical support is an evidence-based strategy shown to improve medication adherence and BP management [8]. In SMBP, a patient uses a BP monitor (BPM) to measure and record their BP readings outside of a clinical environment, usually at home [9]. This strategy has the potential to enhance the quality and accessibility of care for patients with hypertension. The use of SMBP that capitalizes on digitally connected devices allows measurements to be transferred to patients’ electronic health records (EHRs).

Despite SMBP being recommended as a successful strategy for hypertension management, there is a lack of infrastructure to enable proper SMBP transmission of BP measurements [10]. Patients encounter difficulties in sharing SMBP readings with providers outside of typical in-person visits, whereas providers face difficulties in incorporating SMBP data into their clinical workflow to provide timely patient feedback [10]. The digital management of hypertension through remote patient monitoring (RPM) supports SMBP transmission of BP readings. RPM is a technology that allows patients to be monitored outside of traditional clinical settings, such as at home, thereby increasing access to care and lowering health care delivery costs. Digital health and RPM can offer providers a more holistic perspective on their patients’ health and allow patients to take more control over their health.

This Study

Few existing programs integrate SMBP with RPM to interpret BP data, encourage lifestyle changes, and titrate medication [11]. The purpose of this pilot study was to determine the feasibility and acceptability of a digitally based BP self-monitoring program that promotes hypertension self-management and health education among low-income patients. We hypothesized that the program would be highly feasible and acceptable to patients and that at least 50% of the patients would use their BPM at the rate required for the reimbursement of the device’s cost (16 days of measurements in any 30-day period).

Methods

Study Design and Study Population

This study was conducted at Family Health Centers of San Diego (FHCSD), the largest FQHC system in San Diego, California. FHCSD’s primary care clinics are strategically located in federally designated health professional shortage areas and serve medically underserved areas with high proportions of uninsured patients. FHCSD serves >160,000 unduplicated patients annually, the vast majority of whom are low-income individuals and members of a minority population. Approximately 32% are uninsured, 37% are best served in a language other than English, and >55% are Hispanic. The inclusion criteria for the digital health program were as follows: being an FHCSD patient who (1) was aged ≥18 years; (2) spoke English, Spanish, or Arabic; (3) had an appointment with FHCSD within the last 6 months; and (4) had a diagnosis or history of hypertension in their EHR. Exclusion criteria included not having completed an FHCSD Broad Consent form.

We deployed 180 cellularly connected BPMs to patients. The rolling enrollment period began in January 2022 and ended in July 2022. The monitors were deployed in two ways: (1) in person and (2) via the web. The goal was to deploy 90 devices remotely and 90 in person. By July, a total of 179 patients were onboarded through Zoom (Zoom Video Communications, Inc) and 89 (49.7%) of whom were onboarded remotely and 90 in person. The monitors would use their BPM at the rate required for the reimbursement of the device’s cost (16 days of measurements in any 30-day period).
Training and Onboarding

During onboarding, each patient was trained on how to use the BPM and how to interpret BP readings. The patients also received health education on the importance of monitoring and managing hypertension.

The health education content that was provided to the participants included (1) background on SMBP, (2) how to properly prepare for self-measurement (eg, avoiding caffeine, smoking, and exercise from 30 minutes before measuring; waiting at least 30 minutes after a meal; and emptying bladder), (3) proper positioning for self-measurement (eg, feet flat on the floor, legs uncrossed, back straight, and arm supported or palm up), (4) how to use the BPM, (5) how to properly self-measure (eg, resting for 5 minutes before starting; relaxing the body; avoiding conversations, TV, or phones; and taking 2 to 3 measurements 1 minute apart), and (6) how to read and understand SMBP measurements and categories (normal, elevated, stage 1, stage 2, and hypertension crisis). We used the teach-back method to ensure that the patients understood the information provided and to address any additional questions or concerns from the patient.

Measurement

Patients were advised to measure their BP at home twice per day, once in the morning and once in the late afternoon, per the American Heart Association’s (AHA) evidence-based guidelines for home BP monitoring. However, if they were unable to do so, we recommended that patients measure their BP at least every other day for 3 months using the Withings BPM Connect. This device is very user-friendly, is smaller than most BPMs, provides accurate measurements, is cellular enabled (ie, it can be used with or without a smartphone), and displays numerical readings with large LED lights and color-coded indicators [12].

Our recommended measurement plan of measuring at least every other day was based on the Centers for Medicare and Medicaid Services reimbursement requirement of at least 16 days of monitoring over a 30-day period [13]. Specifically, monitoring must occur on at least 16 days over a 30-day period for Current Procedural Terminology (CPT) codes 99453 and 99454 to be billed. CPT code 99453 is valued to reflect clinical staff time that includes instructing a patient about using one or more medical devices, and CPT code 99454 is valued to include the medical device supplied to the patient and programming of the medical device for repeated monitoring. Although measuring every other day during 30-day period equals 15/30 readings, assuming that the patients attempted to measure every day on occasion, this would have got us close to receiving at least 16/30 days of data within a 30-day period.

Surveys

Patients were asked to complete a baseline survey to characterize digital health literacy and hypertension health outcomes, as well as a postprogram survey after 3 months of SMBP to gather feedback on the program and device usability (Multimedia Appendices 1 and 2). Baseline hypertension questions were developed with reference to the AHA and American Medical Association “Lower Your Blood Pressure” questionnaire, which collects information about medications, lifestyle changes, and challenges with managing BP [14]. Baseline digital health questions were developed using the Digital Health Literacy Instrument, which measures familiarity and the ability to operate digital devices [15].

For the postprogram survey, usability questions on the Withings BPM Connect were developed using the System Usability Scale (SUS), a 10-item survey with 5 response options ranging from strongly agree to strongly disagree [16]. This scale may be used to test a wide range of products and services, including hardware, software, mobile devices, websites, and apps [16]. The SUS items were modified to use a simpler language that our patient population could understand, and 2 of the items were removed. Item 6 (“I thought there was too much inconsistency in this system”) and item 8 (“I found the system very cumbersome to use”) were eliminated because of being redundant with other questions and having different meanings when translated into Spanish. In the 10-item SUS survey, scores >68 are considered above average [16]. In this study, the scoring multiplier was adjusted from 100/40 to 100/32 to accommodate the reduced number of questions [17].

Clinical Support

To support healthy lifestyles and behavior change, educational booklets were created with colorful infographics clearly illustrating important BP concepts (Multimedia Appendix 3). The booklets included step-by-step instructions for taking accurate BP readings, information on readings and BP categories, and tips for success (eg, eat smart, move more, manage weight, do not smoke, and sleep well). The booklets were created in 3 languages: English, Spanish, and Arabic. Patients who were onboarded in person received a physical copy of the booklet, whereas patients who were onboarded remotely received a PDF version of the booklet through email.

To engage patients at home, an SMS text messaging campaign was offered (Multimedia Appendix 4). For 7 weeks, the participating patients received 1 SMS text message delivered at 9 AM daily. The SMS text messages were designed to remind patients to check their BP and to promote a healthier diet, physical activity, more sleep, stress awareness, and weight loss. All SMS text message content was evidence based and had previously been shown to stimulate behavior change [18,19].

After 3 months, when the postprogram survey was completed, the patients were asked whether they wished to continue using their BPM for another 3 months. We reassured them that we would continue to monitor their results and alert their providers of any elevated readings. If the patients agreed to continue using their monitor, they were sent monetary incentives (US $25 gift cards) via mail to thank them for continuing to use their devices and to encourage them to continue self-monitoring at home, with the goal of increasing measurement adherence even after program completion.

Withings RPM

The Withings RPM was used to access and review patient data. This platform allows health care teams to create a customized measuring plan and collaborate with other professionals to manage multiple patients. The Withings RPM includes a feature that provides automated alerts for health care teams to review
whenever a patient’s BP level is outside of the normal range. Measurement results (date, time, BP, and heart rate) were automatically sent to the platform, and FHCSD’s team of digital health specialists monitored the patients’ readings and alerts daily. If a patient continuously had elevated BP readings, their primary care provider was contacted via email with the suggestion to contact the patient and schedule a follow-up BP check visit. The protocol consisted of the following:

- If a patient triggered a yellow alert ($\geq 160$ mm Hg systolic BP [SBP] or $\geq 100$ mm Hg diastolic BP [DBP]), the team monitored their measurements for the next 7 to 10 days. If the patient consistently had yellow alerts for those 7 to 10 days, the team then proceeded to send a message to the patient’s provider within 24 to 48 hours.

- If a patient triggered a red alert ($\geq 180$ mm Hg SBP or $\geq 120$ mm Hg DBP), the team sent a message to their provider immediately within 24 to 48 hours. The team then continued to monitor their measurements for the next 7 to 10 days. If the patient consistently had red alerts for those 7 to 10 days, the team then proceeded to send an additional message to the patient’s provider within 24 to 48 hours.

The following are examples of the emails that were sent to providers.

This email was sent to providers via EHRs for consistent yellow alerts:

```
Hello, your patient has agreed to receive a digitally connected blood pressure monitor. We are monitoring patients’ readings and notifying their provider if readings are elevated. Your patient has been consistently having elevated blood pressure readings within the past week (>160 mmHg SBP or >100 mmHg DBP). We think that you or your team could contact them to schedule a follow-up blood pressure check appointment if you wish. Please feel free to reach out to us at digitalhealth@fhcsd.org if you have any questions. Thank you.
```

This email was sent to providers via EHRs for the first red alert:

```
Hello, your patient has agreed to receive a digitally connected blood pressure monitor. We are monitoring patients’ readings and notifying their provider if readings are too high. Your patient currently has out of range blood pressure readings (>180 mmHg SBP or >120 mmHg DBP). We think that you or your team could contact them to schedule a follow-up blood pressure check appointment if you so wish. Please feel free to reach out to us at digitalhealth@fhcsd.org if you have any questions. Thank you.
```

This email was sent to providers via EHRs for consistent red alerts:

```
Hello, this is a follow-up from my last email. Your patient has been consistently continuing to have out of range blood pressure readings within the past week (>180 mmHg SBP or >120 mmHg DBP). I am just sharing this again in case you or your team would like to schedule a follow-up blood pressure check appointment with them. Please feel free to reach out to us at digitalhealth@fhcsd.org if you have any questions. Thank you.
```

---

**Statistical Analysis**

All data processing, analyses, and calculations were performed in R (version 4.1.2 Bird Hippie; R Foundation for Statistical Computing). Raw data from the BPMs were processed and cleaned, and patients who never activated their monitors were removed from the data set. Descriptive statistics were used to describe the demographic characteristics of the study population. Responses to the baseline digital health literacy survey questions were analyzed using descriptive statistics. Calculations related to compliance, use, mean BP, and BP control status were performed. Descriptive statistics were used to analyze the responses to the postprogram survey.

For the time-to-event analysis, a Kaplan-Meier analysis was conducted to compare time to compliance between 2 definitions of compliance: a strict definition that defined compliance as measuring BP on at least 16 days during any rolling 30-day window versus a lenient definition that defined compliance as measuring BP on at least 1 day per week for any 4 consecutive weeks [20]. For both definitions, measures on the day of onboarding were excluded. Each noncompliant patient was censored after their total number of days in the study. A log-rank test was performed to determine whether the difference in time to compliance between the definitions was statistically significant. An unadjusted (crude) Cox proportional hazards regression model was run to assess the time-to-event distributions by compliance definition group [21]. As the patient population was the same for both definitions, no additional variables (eg, demographic or clinical) could be added to the model; all of the effect would have been based on the definition. The proportional hazards assumption was tested using the `survival` package of R [22,23].

For the comparison of BP measurements taken before and after the intervention, BP data from FHCSD’s EHR system were obtained for a 12-month period prior to when each participant received the intervention. Mean preintervention SBP and DBP were calculated for each patient (except for patients with only 1 BP reading). For each patient, the preintervention mean was subtracted from the postintervention mean to calculate the changes in SBP and DBP. The mean changes in SBP and DBP across patients were calculated. The mean SBP and DBP before the intervention and after the intervention were also calculated across patients. Paired sample t tests (2-tailed) were performed to assess the changes in SBP and DBP from before to after the intervention. Statistical significance was defined as $P<.05$ for all analyses.

**Qualitative Survey Analysis**

The participants were individually surveyed at the end of the program about their experience with the program and asked to numerically rate various aspects of the devices. At the end of this survey, the participants were asked open-ended questions to collect otherwise undisputed feedback. Upon the completion of the study, these responses were reviewed, and broad, mutually exclusive themes were identified.
Ethics Approval

This no more than minimal risk feasibility and acceptability study was conducted as a component of quality improvement activities at FHCSD, a covered entity [24], and there is no requirement for such activities to undergo review by an Institutional Review Board [25]. However, all patients aged >18 years must complete and sign FHCSD’s Broad Consent form to receive treatment. The Broad Consent form includes a specific authorization for the use of deidentified health information for population health and quality improvement studies. We have only included patients who have an up-to-date Broad Consent form within their EHR.

Results

Baseline Characteristics

Between January 2022 and July 2022, a total of 179 patients received BPMs at FHCSD. Of the 179 patients, 89 (49.7%) received a BPM and health education from a digital health specialist in person, and 90 (50.3%) received a BPM by mail and remote health education via Zoom. The demographic characteristics of the study population are shown in Table 1. Overall, the patients had a mean age of 55.1 (SD 12.0) years, and 57.5% (103/179) were women. Most patients were Hispanic (132/179, 73.7%) and reported Spanish as their preferred language (112/179, 62.6%). The most prevalent educational level was less than high school (74/179, 41.3%), followed by high school (63/179, 35.2%), and most participants (103/179, 57.5%) were unemployed.

All 179 patients were offered the SMS text messaging campaign; 120 (67%) opted into this campaign, and 59 (33%) declined. Of the 120 patients who opted into the SMS text message campaign, 45 (37.5%) were English speakers, and 75 (62.5%) were Spanish speakers.

At baseline, all 179 patients completed the high BP and digital health literacy survey. Table 2 shows the study population’s baseline digital health literacy question responses. Most patients reported owning a smartphone (141/179, 78.8%) and had access to the internet or Wi-Fi at home (132/179, 73.7%), whereas 59.2% (106/179) of the participants reported that they did not have a computer, a laptop, or an iPad (Apple, Inc) at home.
<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>103 (57.5)</td>
</tr>
<tr>
<td>Men</td>
<td>76 (42.5)</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
</tr>
<tr>
<td>Values, mean (SD)</td>
<td>55.1 (12.0)</td>
</tr>
<tr>
<td>Values, median (range)</td>
<td>56.0 (26.0-84.0)</td>
</tr>
<tr>
<td><strong>Age group (years), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>18-29</td>
<td>4 (2.2)</td>
</tr>
<tr>
<td>30-39</td>
<td>15 (8.4)</td>
</tr>
<tr>
<td>40-49</td>
<td>33 (18.4)</td>
</tr>
<tr>
<td>50-59</td>
<td>58 (32.4)</td>
</tr>
<tr>
<td>≥60</td>
<td>69 (38.5)</td>
</tr>
<tr>
<td><strong>Race, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>4 (2.2)</td>
</tr>
<tr>
<td>Black</td>
<td>20 (11.2)</td>
</tr>
<tr>
<td>White</td>
<td>151 (84.4)</td>
</tr>
<tr>
<td>Multiracial</td>
<td>4 (2.2)</td>
</tr>
<tr>
<td><strong>Ethnicity, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>132 (73.7)</td>
</tr>
<tr>
<td>Non-Hispanic</td>
<td>46 (25.7)</td>
</tr>
<tr>
<td>Unknown</td>
<td>1 (0.6)</td>
</tr>
<tr>
<td><strong>Education level, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Less than high school</td>
<td>74 (41.3)</td>
</tr>
<tr>
<td>High school</td>
<td>63 (35.2)</td>
</tr>
<tr>
<td>College graduate</td>
<td>18 (10.1)</td>
</tr>
<tr>
<td>Postgraduate</td>
<td>4 (2.2)</td>
</tr>
<tr>
<td>Unknown</td>
<td>20 (11.2)</td>
</tr>
<tr>
<td><strong>Employment status, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>63 (35.2)</td>
</tr>
<tr>
<td>Unemployed</td>
<td>103 (57.5)</td>
</tr>
<tr>
<td>Retired</td>
<td>13 (7.3)</td>
</tr>
<tr>
<td><strong>Preferred language, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>61 (34.1)</td>
</tr>
<tr>
<td>Spanish</td>
<td>112 (62.6)</td>
</tr>
<tr>
<td>Other</td>
<td>6 (3.4)</td>
</tr>
<tr>
<td><strong>Deployment type, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>In person</td>
<td>89 (49.7)</td>
</tr>
<tr>
<td>Remote</td>
<td>90 (50.3)</td>
</tr>
</tbody>
</table>
Table 2. Baseline digital health literacy question responses (N=179).

<table>
<thead>
<tr>
<th>Question and response</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do you have a smartphone?</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>141 (78.8)</td>
</tr>
<tr>
<td>No</td>
<td>38 (21.2)</td>
</tr>
<tr>
<td>Do you have a computer, laptop, or iPad at home?</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>73 (40.8)</td>
</tr>
<tr>
<td>No</td>
<td>106 (59.2)</td>
</tr>
<tr>
<td>Do you have internet and/or Wi-Fi at home?</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>132 (73.7)</td>
</tr>
<tr>
<td>No</td>
<td>46 (25.7)</td>
</tr>
<tr>
<td>Missing</td>
<td>1 (0.6)</td>
</tr>
</tbody>
</table>

Compliance

Of the 179 patients who received a BPM, 168 (93.9%) activated their devices, and 11 (6.1%) did not. Approximately 6 months after rolling enrollment began in January 2022, a total of 26.8% (45/168) of the patients were compliant with the recommended measuring plan of 16 out of 30 days, the measurement rate that would result in device cost reimbursement. Most patients (123/168, 73.2%) did not achieve this target.

Use of BP Monitors

Figure 1 illustrates the percentage of patients who used their BPMs for >26 weeks (approximately 6 months) after rolling enrollment began in January 2022. On the day of onboarding, the digital health specialist took 1 measurement for each patient, resulting in a 100% (n=168) use rate for week 1 (week 1 was defined as the first week of BP measurement for each patient). The use of the BPMs decreased over time, with 53.6% (90/168) of the patients using their monitor at week 2, a total of 42.3% (71/168) using their monitor at week 4, and 35.7% (60/168) using their monitor at week 6. At week 11, a total of 25% (42/168) of the patients were using their monitor; at week 26, the use rate was 4.2% (7/168).
BP Control

Figure 2 illustrates the mean SBP and DBP across participants by week, with the number of participants decreasing over time (n=168 at week 1; n=7 at week 26). Both measures were consistent over time. Overall, the mean SBP was 136.2 (SD 19.6) mm Hg, and the mean DBP was 82.2 (SD 13.6) mm Hg. Over the entire study period, there were 913 yellow alerts (15.4% of all 5935 readings) and 75 red alerts (1.3% of all 5935 readings).

Preintervention BP data from EHRs were available for 154 (86%) out of 179 patients. The mean change in mean SBP from before to after the intervention was +2.62 (SE 1.26) mm Hg. The mean change in mean DBP from before to after the intervention was +3.31 (SE 0.71) mm Hg. Overall, 42.2% (65/154) of the patients had a decrease in SBP following the intervention, and 57.8% (89/154) of the patients had an increase in SBP following the intervention, whereas 30.5% (47/154) of the patients had a decrease in DBP following the intervention, and 69.5% (107/154) of the patients had an increase in DBP following the intervention. The mean SBP and DBP before and after the intervention are shown in Figure 3. The paired sample *t* test (2-tailed) conducted to assess the change in SBP from before to after the intervention produced the following results (*t*<sub>153</sub>=2.0762; *P*=.04). The paired sample *t* test conducted to assess the change in DBP from before to after the intervention produced the following results (*t*<sub>153</sub>=4.6702; *P*<.001). For both *t* tests, the *P* value was <.05, indicating that there was a statistically significant increase in both SBP and DBP after the intervention compared to before the intervention.

Table 3 shows the participants’ BP control status at the first, last, and any measures—37.5% (63/168) of the patients were in control at their first BP measure, and 48.8% (82/168) were in control at their last measure. Although 83.3% (140/168) of the patients had at least 1 measure in control, 16.7% (28/168) did not achieve control at any measure.

Figure 1. Use of the blood pressure monitors over time (n=168). Patients participated in the program for different durations, with some participating for <26 weeks.
Figure 2. Mean systolic blood pressure (SBP) and diastolic blood pressure (DBP) by week (n=168). Shaded regions represent 95% CI for weekly means. BP: blood pressure.
Figure 3. Mean systolic blood pressure (SBP) and diastolic blood pressure (DBP) before and after the intervention (n=154). BP: blood pressure.

Table 3. Blood pressure control status at first, last, and any measures (n=168).

<table>
<thead>
<tr>
<th></th>
<th>Uncontrolled, n (%)</th>
<th>Controlled, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>First measure</td>
<td>105 (62.5)</td>
<td>63 (37.5)</td>
</tr>
<tr>
<td>Last measure</td>
<td>86 (51.2)</td>
<td>82 (48.8)</td>
</tr>
<tr>
<td>Any measure</td>
<td>28 (16.7)</td>
<td>140 (83.3)</td>
</tr>
</tbody>
</table>

Time to Compliance

Time to compliance was compared between 2 definitions of compliance: a strict definition that defined compliance as measuring BP on at least 16 days during any rolling 30-day window and a lenient definition that defined compliance as measuring BP on at least 1 day per week for any 4 consecutive weeks. On the basis of the strict definition, 25.6% (43/168) of patients achieved compliance, and 74.4% (125/168) of patients did not achieve compliance. For the strict definition, the mean time to compliance was 119 days, and the median time to compliance was 130 (range 29-207) days. On day 29, the first opportunity, 32 (19%) of 168 patients met the strict definition of compliance. For the lenient definition, 42.3% (71/168) of patients achieved compliance, and 57.7% (97/168) of patients did not achieve compliance. For the lenient definition, the mean time to compliance was 103 days, and the median time to compliance was 95 (range 27-207) days. On day 27, the first opportunity, 26.2% (44/168) of patients met the lenient definition of compliance. Figure 4 shows the time to compliance for each definition. The log-rank test showed that the difference in time to compliance between the compliance definitions was statistically significant (P<.001). The unadjusted (crude) Cox proportional hazards regression model showed that there was a statistically significant difference in time to event between the compliance definitions (hazard ratio=0.51; P<.001, for the strict definition compared with the lenient definition; test for the proportional hazards assumption had a P value of .007).
Feasibility and Acceptability

At follow-up, 83.2% (149/179) of patients completed the postprogram survey. Of the 179 patients who completed the baseline survey, 30 (16.8%) were lost to follow-up (≥4 call attempts with no answer or phone number disconnected). The postprogram survey responses, presented in Table 4, indicate a high level of feasibility and acceptability. When asked to rate the BPM, over three-quarters (113/149, 75.8%) of the participants selected “very good” (55/149, 36.9%) or “good” (58/149, 38.9%). Furthermore, 91.9% (137/149) of the participants reported that they would recommend the BPM to others. A very small percentage of participants (3/149, 2%) rated the BPM as either “bad” or “very bad,” and 71.8% (107/149) of the participants reported that they would be willing to continue using the BPM for another 3 months.

In the postprogram survey, the participants were asked 8 questions from the SUS. The mean score was 62.7 (SD 16.6), and the median score was 68.8 (Table 5). This corresponds to a qualitative interpretation of the patients scoring Withings BPM Connect as average in regard to effectiveness, efficiency, and satisfaction. The patients were also prompted to provide open-ended feedback using 2 questions: “Do you think the blood pressure monitor is helping you better manage your hypertension?” and “Please feel free to share with us any other feedback you might have.” In answering these questions, many patients provided context for their decreased or discontinued use of the BPM. Table 6 details the prevalence of the most common responses. The most frequent barriers to device use were related to inconvenience due to medical, professional, or personal obligations. Many patients also reported challenges in operating the device, concerns about accuracy, the lack of medical need, forgetfulness, lost or stolen devices, and fear as reasons for not regularly using their BPM.

Table 4. Postprogram survey responses (n=149).

<table>
<thead>
<tr>
<th>Question and response</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>How would you rate the blood pressure monitor?</td>
<td></td>
</tr>
<tr>
<td>Very good</td>
<td>55 (36.9)</td>
</tr>
<tr>
<td>Good</td>
<td>58 (38.9)</td>
</tr>
<tr>
<td>Neutral</td>
<td>33 (22.1)</td>
</tr>
<tr>
<td>Bad</td>
<td>2 (1.3)</td>
</tr>
<tr>
<td>Very bad</td>
<td>1 (0.7)</td>
</tr>
<tr>
<td>Would you recommend the blood pressure monitor to other patients?</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>137 (91.9)</td>
</tr>
<tr>
<td>No</td>
<td>12 (8.1)</td>
</tr>
<tr>
<td>Are you willing to continue using the blood pressure monitor for another 3 months?</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>107 (71.8)</td>
</tr>
<tr>
<td>No</td>
<td>42 (28.2)</td>
</tr>
<tr>
<td>Calculated SUS score</td>
<td>Values, mean (SD)</td>
</tr>
<tr>
<td>----------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Values, median (range)</td>
<td></td>
</tr>
</tbody>
</table>

I think that I would like to use the Withings BPM frequently, n (%)

<table>
<thead>
<tr>
<th></th>
<th>Strongly agree</th>
<th>32 (21.5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agree</td>
<td>47 (31.5)</td>
<td></td>
</tr>
<tr>
<td>Neutral</td>
<td>45 (30.2)</td>
<td></td>
</tr>
<tr>
<td>Disagree</td>
<td>23 (15.4)</td>
<td></td>
</tr>
<tr>
<td>Strongly disagree</td>
<td>1 (0.7)</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>1 (0.7)</td>
<td></td>
</tr>
</tbody>
</table>

I found the Withings BPM complicated to use, n (%)

<table>
<thead>
<tr>
<th></th>
<th>Strongly agree</th>
<th>8 (5.4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agree</td>
<td>16 (10.7)</td>
<td></td>
</tr>
<tr>
<td>Neutral</td>
<td>25 (16.8)</td>
<td></td>
</tr>
<tr>
<td>Disagree</td>
<td>73 (49)</td>
<td></td>
</tr>
<tr>
<td>Strongly disagree</td>
<td>27 (18.1)</td>
<td></td>
</tr>
</tbody>
</table>

I thought the Withings BPM was easy to use, n (%)

<table>
<thead>
<tr>
<th></th>
<th>Strongly agree</th>
<th>31 (20.8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agree</td>
<td>73 (49)</td>
<td></td>
</tr>
<tr>
<td>Neutral</td>
<td>33 (22.1)</td>
<td></td>
</tr>
<tr>
<td>Disagree</td>
<td>10 (6.7)</td>
<td></td>
</tr>
<tr>
<td>Strongly disagree</td>
<td>2 (1.3)</td>
<td></td>
</tr>
</tbody>
</table>

I think that I would need the support of a digital health specialist to use the Withings BPM, n (%)

<table>
<thead>
<tr>
<th></th>
<th>Strongly agree</th>
<th>23 (15.4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agree</td>
<td>22 (14.8)</td>
<td></td>
</tr>
<tr>
<td>Neutral</td>
<td>34 (22.8)</td>
<td></td>
</tr>
<tr>
<td>Disagree</td>
<td>67 (45)</td>
<td></td>
</tr>
<tr>
<td>Strongly disagree</td>
<td>3 (2)</td>
<td></td>
</tr>
</tbody>
</table>

I found the Withings BPM's measurements easy to read, n (%)

<table>
<thead>
<tr>
<th></th>
<th>Strongly agree</th>
<th>51 (34.2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agree</td>
<td>74 (49.7)</td>
<td></td>
</tr>
<tr>
<td>Neutral</td>
<td>21 (14.1)</td>
<td></td>
</tr>
<tr>
<td>Disagree</td>
<td>3 (2)</td>
<td></td>
</tr>
</tbody>
</table>

I would imagine that most people would learn to use the Withings BPM very quickly, n (%)

<table>
<thead>
<tr>
<th></th>
<th>Strongly agree</th>
<th>3 (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agree</td>
<td>89 (59.7)</td>
<td></td>
</tr>
<tr>
<td>Neutral</td>
<td>39 (26.2)</td>
<td></td>
</tr>
<tr>
<td>Disagree</td>
<td>18 (12.1)</td>
<td></td>
</tr>
</tbody>
</table>

I felt very confident using the Withings BPM, n (%)

<table>
<thead>
<tr>
<th></th>
<th>Strongly agree</th>
<th>11 (7.4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agree</td>
<td>77 (51.7)</td>
<td></td>
</tr>
<tr>
<td>Neutral</td>
<td>37 (24.8)</td>
<td></td>
</tr>
</tbody>
</table>
Table 6. Feedback from the patients on the usability of the blood pressure monitor (n=149).

<table>
<thead>
<tr>
<th>Reason for decreased or discontinued use</th>
<th>Prevalence, n (%)</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inconvenience in using the device</td>
<td>41 (27.5)</td>
<td>• Prioritizing another health condition</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Busy schedule due to personal or professional obligations</td>
</tr>
<tr>
<td>Challenge in operating the device</td>
<td>36 (24.2)</td>
<td>• Cannot use the device while alone, as help from family is required to use the device</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Physical discomfort in using the device</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Technical issue, namely device malfunction or confusion in operating the device</td>
</tr>
<tr>
<td>Concerns about accuracy</td>
<td>16 (10.7)</td>
<td>• Results do not match the readings of other blood pressure monitoring devices.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Results do not reflect the patient’s expectations.</td>
</tr>
<tr>
<td>Belief that use was not medically necessary</td>
<td>15 (10.1)</td>
<td>—a</td>
</tr>
<tr>
<td>Forgetting to use the device</td>
<td>14 (9.4)</td>
<td>—</td>
</tr>
<tr>
<td>Lost or stolen device or charger</td>
<td>10 (6.7)</td>
<td>—</td>
</tr>
<tr>
<td>Fear of viewing worrisome results</td>
<td>8 (5.4)</td>
<td>—</td>
</tr>
</tbody>
</table>

*aNot available.

Discussion

Principal Findings

Results of the baseline digital health literacy survey revealed that most patients owned a smartphone (141/179, 78.8%) and had access to the internet or Wi-Fi at home (132/179, 73.7%), but less than half (73/179, 40.8%) of the participants had a computer, a laptop, or an iPad at home. Before implementing this program, we hypothesized that at least 50% of the patients would use their monitor at the rate required for the reimbursement of the cost of the device (16 days of measurements in any 30-day period). Approximately 6 months after the beginning of rolling enrollment in January, use and compliance data did not support this hypothesis. Only 27% (45/168) of the patients complied with the measurement plan that would result in device cost reimbursement, whereas 73% (123/168) of the patients did not achieve this target. The fact that such a low proportion of patients complied with the recommended measurement rate that would result in device cost reimbursement suggests that the current criteria for reimbursement may be too stringent and, therefore, inappropriate in lower resource settings. Furthermore, if a patient’s BP is under control, it is instinctual to reduce the frequency of home measures. Therefore, the likelihood of compliance with a reimbursable measurement threshold may decrease with increased BP control. It is critical for those creating the logic behind the financial support of RPM programs to consider this potential.

The use of BPMs declined over time, with use rates of 53.6% (90/168) at week two, 42.3% (71/168) at week four, 35.7% (60/168) at week six, 25% (42/168) at week 11, and 4.2% (7/168) at week 26. There was a slight but statistically significant increase in both SBP and DBP following the intervention. Only 25.6% (43/168) of the patients achieved the strict definition of compliance (measuring at least 16 days out of any 30 consecutive days), whereas 42.3% (71/168 of the patients achieved the lenient definition (measuring at least 1 day per week for 4 consecutive weeks). The difference in time to compliance between these compliance definitions was statistically significant ($P<.001$). With respect to BP control status, 37.5% (63/168) of the patients were in control at their first measurement, 48.8% (82/168) were in control at their last measurement, and 83.3% (140/168) had at least 1 measurement in control. In total, 16.7% (28/168) of the patients did not achieve control at any measurement. This observation has important implications for the definitions of clinical quality.
Specifically, the Uniform Data System and Health Care Effectiveness Data and Information Set define hypertension control as SBP <140 mm Hg and DBP <90 mm Hg [26,27]. Furthermore, during a calendar year, every patient aged 18 to 85 years with a hypertension diagnosis should have a controlled BP reading documented during a qualified medical visit. When assessing the compliance with the Uniform Data System or Health Care Effectiveness Data and Information Set measure of hypertension control, only the last BP reading in a patient’s medical record in a calendar year is considered [26,27]. In the context of RPM, a new BP reading can appear in a patient’s medical record on a daily basis. This raises fundamental questions about the operational definition of quality when it comes to hypertension control. Our numbers indicated that the potential compliance differs by as much as 46%, depending on what is chosen as the last observation. With this in mind, researchers, scientists, clinicians, and policy makers must come together to further explore the implications of RPM and its impact on clinical quality.

We also hypothesized that our digital health program would have high feasibility and acceptability among our study population. At follow-up, the postprogram survey results corroborated this hypothesis, with most participants (113/149, 75.8%) rating the BPM as either “very good” or “good” and a very small percentage of participants (3/149, 2%) rating the BPM as either “bad” or “very bad.” Most participants reported that they would recommend the BPM to others (137/149, 91.9%) and that they would be willing to continue using the BPM for another 3 months (107/149, 71.8%). These results indicate a high level of acceptability, which is interesting given the low use and compliance rates observed over time. Although the patients recommended these devices to others and positively rated their experience, the additional feedback offered at the end of the survey revealed several barriers that made it challenging for them to use their BPM. The patients reported experiencing issues with convenience, technical challenges, and concerns about accuracy. Although patients seemed to recognize the benefits of the BPM, these issues may have outweighed their interest in continued regular use. This feedback provides an interesting context for the disparity between our positive survey results and the low use and compliance rates.

**Comparison With Prior Work**

The effectiveness of RPM programs at improving the management of chronic diseases, including hypertension, has been established. A 2022 overview of recent systematic reviews used the Grading of Recommendations Assessment, Development, and Evaluation approach to evaluate the certainty of the evidence among randomized controlled trials using RPM in adult patients with hypertension, diabetes, or both [28]. The findings suggested that RPM likely caused a small decrease in SBP, although it was unclear whether this decrease was clinically meaningful [28]. The Community Preventive Services Task Force reported finding sufficient evidence of the effectiveness of SMBP interventions when used alone but strong evidence of the effectiveness of SMBP interventions when combined with additional support (eg, patient education, counseling, or web-based support) [29].

The effectiveness of RPM programs at improving hypertension management, specifically in underserved patient populations, has previously been explored, with promising findings. A study assessing the effectiveness of SMBP among medically underserved, low-income Black and Hispanic patients with hypertension reported a decrease in both SBP and DBP at follow-up [30]. A pilot study implementing the Measure Accurately, Act Rapidly, and Partner With Patients evidence-based protocol to address hypertension among a medically underserved patient population identified a clinically and statistically significant improvement in hypertension control 6 months after the intervention [31]. A study in Hispanic adults with uncontrolled hypertension analyzed the efficacy of a culturally tailored, smartphone-enabled self-management program that included SMBP and reported clinically and statistically significant reductions in SBP following the intervention [32]. A pilot study assessing the use of a tailored mobile health intervention in Black patients with hypertension and type 2 diabetes reported statistically significant improvements in SBP [33]. A study of an intensive RPM program focused on hypertension and deployed in an FQHC serving low-income Asian American patients reported that 96% of the patients improved their BP control following the intervention [34]. By contrast, this study found little change over time with respect to SBP or DBP, suggesting that patients may require an increased level of support while undergoing RPM to achieve meaningful reductions in BP.

Studies examining the feasibility and acceptability of RPM programs targeting hypertension in underserved patient populations have identified both positive and negative aspects of patients’ experiences. A study investigating knowledge of hypertension, engagement in care, and attitudes toward and experiences with SMBP in patients from 9 community health centers found that most patients (85%) reported having positive experiences with SMBP and that patients’ engagement with care increased significantly after SMBP [35]. The authors suggested that patients’ positive experiences were attributable to the fact that they were provided with education, training, and support while engaging in SMBP [35]. A mixed methods study conducted in Hmong and Latino adults with hypertension at an FQHC to examine patients’ perspectives about a mobile health–based care model including RPM found that patients found the program useful, especially if they were provided assistance with navigating technological challenges [36]. Sharing their BP data with the clinic felt empowering to some patients but entrapping to others [36]. A pilot study examining the feasibility and effectiveness of training high school students as health technology coaches to help medically underserved patients with hypertension found that, compared with the BPM-only group, the BPM-plus-health-coaching group had a higher frequency of self-monitoring, higher engagement and satisfaction, and better self-reported BP [37]. Patients expressed concerns regarding the inconsistency of results and reported that their health coaches helped them troubleshoot technical difficulties [37]. A systematic review of qualitative studies assessing patients’ experiences of RPM for chronic disease management found that RPM improved disease-specific knowledge, self-management, and decision-making and initiated earlier clinical assessment [38]. Patients’ concerns with RPM...
included a loss of interpersonal contact and increased personal responsibility [38]. A study of an intensive RPM program in an FQHC serving low-income Asian American patients identified low digital health literacy and a lack of in-language digital training as barriers to successful RPM [34].

The barriers to successful RPM uncovered in previous studies are similar to those identified in this study, in which patients expressed concerns about device accuracy and convenience and reported experiencing technical challenges with using their BPM. The body of evidence suggests that patient education and support are integral to successful RPM programs in low-resource settings. Therefore, the barriers identified in this study may be overcome by incorporating more intensive patient education and increasing patients’ access to personalized support from program staff. Future research should compare the different forms and intensities of patient education and support to discover which are most effective in underserved patient populations.

**Strengths and Limitations**

This study has several key strengths. It is unique in that it examined an SMBP program that included training, education, and outreach and was implemented at an FQHC among underserved patients, many of whom were Latino and Spanish speaking. Our setting and population render our findings highly generalizable to other FQHCs; our findings can be leveraged to support further research, build capacity, and improve the digital health infrastructure of other FQHCs. The digital health program at FHCSD was designed under the premise that patient-provider relationships and health care professional support are essential for SMBP to improve hypertension outcomes. As such, the program was supported by digital health specialists who received direct training from the AHA on how to develop, implement, and manage a high-quality digital health program focused on SMBP. Our program included both one-on-one training on how to use a BPM and health education both in person at clinics and via telehealth encounters. Patients were also offered technical assistance with using their BPM.

This study also has important limitations. Although we assessed changes in SBP and DBP over time, our study design did not enable us to analyze how different factors (eg, demographic or clinical factors) influenced BP control status. Future studies with more robust designs should directly examine which patient characteristics are associated with BP control status. In addition, our data did not permit us to analyze which forms and intensities of patient education and support were most impactful for patients while engaging in RPM. Many patients rated the BPM favorably, but the low use rates suggest that the patients experienced challenges with using the BPM.

**Conclusions**

This study demonstrated the acceptability of a simple, low-cost program for monitoring BP among patients at an FQHC. However, few patients were able to use the BPM at the rate that would result in device cost reimbursement. Such programs may not be financially feasible at scale if reimbursement does not occur. Given that RPM programs show promise in the FQHC setting, future research should focus on evaluating plans for sustaining costs among low-income, underserved patients so that this population is better supported in managing hypertension.

**Acknowledgments**

First and foremost, the authors would like to thank their patients for their time, trust, and perspectives. The authors are grateful to the various staff members and administrative leadership at Family Health Centers of San Diego for providing in-kind support for the development of this study. They are also grateful to Withings and the San Diego division of the American Heart Association for their donation of the blood pressure monitors used in our research. Specifically, they would like to thank Chelsea Walczak Vircks, Senior Director at the American Heart Association, for her advice and support of this project.

The study was made possible by funding from the Conrad Prebys Foundation and the National Institutes of Health (National Heart, Lung, and Blood Institute grant K23HL148530).

**Conflicts of Interest**

None declared.

**Multimedia Appendix 1**

Hypertension and digital health literacy baseline survey template.
[DOCX File, 39 KB - humanfactors_v10i1e46313_app1.docx]

**Multimedia Appendix 2**

Template of the 3-month-postprogram survey.
[DOCX File, 27 KB - humanfactors_v10i1e46313_app2.docx]

**Multimedia Appendix 3**

Health education booklets provided to patients during training and onboarding. We collaborated with our local American Heart Association (AHA) branch to create these booklets for patients.
[PDF File (Adobe PDF File), 9289 KB - humanfactors_v10i1e46313_app3.pdf]
Multimedia Appendix 4
Digital health SMS text message campaign. These are the hypertension-related SMS text messages that were sent to the patients who opted into the SMS campaign.

[DOCX File .44 KB - humanfactors_v10i1e46313_app4.docx ]

References
14. Partnering with your doctor to treat high blood pressure. American Heart Association. URL: https://www.heart.org/en/health-topics/high-blood-pressure/changes-you-can-make-to-manage-high-blood-pressure/partnering-with-your-doctor-to-treat-high-blood-pressure [accessed 2022-12-05]
15. van der Vaart R, Drossaert C. Development of the digital health literacy instrument: measuring a broad spectrum of health literacy-topics/high-blood-pressure/changes-you-can-make-to-manage-high-blood-pressure/partnering-with-your-doctor-to-treat-high-blood-pressure/ [accessed 2022-12-05]
17. Lewis JR, Sauro J. Can I leave this one out? The effect of dropping an item from the SUS. J Usability Stud 2017;13(1):38-46 [FREE Full text]


Abbreviations

AHA: American Heart Association
BP: blood pressure
BPM: blood pressure monitor
DBP: diastolic blood pressure
EHR: electronic health record
FHCSD: Family Health Centers of San Diego
FQHC: federally qualified health center
RPV: remote patient monitoring
SBP: systolic blood pressure
SMBP: self-measured blood pressure
SUS: System Usability Scale
Digitally Based Blood Pressure Self-Monitoring Program That Promotes Hypertension Self-Management and Health Education Among Patients With Low-Income: Usability Study


URL: https://humanfactors.jmir.org/2023/1/e46313

doi:10.2196/46313

PMID:37486745
Operational Implementation of Remote Patient Monitoring Within a Large Ambulatory Health System: Multimethod Qualitative Case Study

Katharine Lawrence1,2*, MD; Nina Singh1*, BS; Zoe Jonassen1*, MA; Lisa L Groom3*, RN; Veronica Alfaro Arias2*, MPS; Soumik Mandal1*, PhD; Antoinette Schoenthaler1*, PhD; Devin Mann1,2*, MD; Oded Nov4*, PhD; Graham Dove5*, PhD

1Department of Population Health, New York University Grossman School of Medicine, New York, NY, United States
2Medical Center Information Technology, NYU Langone Health, New York, NY, United States
3Rory Meyers College of Nursing, New York University, New York, NY, United States
4Tandon School of Engineering, New York University, New York, NY, United States
*all authors contributed equally

Corresponding Author:
Katharine Lawrence, MD
Department of Population Health
New York University Grossman School of Medicine
227 E 30th St
6th Fl
New York, NY, 10016
United States
Phone: 1 646 301 3400
Email: katharine.lawrence@nyulangone.org

Abstract

Background: Remote patient monitoring (RPM) technologies can support patients living with chronic conditions through self-monitoring of physiological measures and enhance clinicians’ diagnostic and treatment decisions. However, to date, large-scale pragmatic RPM implementation within health systems has been limited, and understanding of the impacts of RPM technologies on clinical workflows and care experience is lacking.

Objective: In this study, we evaluate the early implementation of operational RPM initiatives for chronic disease management within the ambulatory network of an academic medical center in New York City, focusing on the experiences of “early adopter” clinicians and patients.

Methods: Using a multimethod qualitative approach, we conducted (1) interviews with 13 clinicians across 9 specialties considered as early adopters and supporters of RPM and (2) speculative design sessions exploring the future of RPM in clinical care with 21 patients and patient representatives, to better understand experiences, preferences, and expectations of pragmatic RPM use for health care delivery.

Results: We identified themes relevant to RPM implementation within the following areas: (1) data collection and practices, including impacts of taking real-world measures and issues of data sharing, security, and privacy; (2) proactive and preventive care, including proactive and preventive monitoring, and proactive interventions and support; and (3) health disparities and equity, including tailored and flexible care and implicit bias. We also identified evidence for mitigation and support to address challenges in each of these areas.

Conclusions: This study highlights the unique contexts, perceptions, and challenges regarding the deployment of RPM in clinical practice, including its potential implications for clinical workflows and work experiences. Based on these findings, we offer implementation and design recommendations for health systems interested in deploying RPM-enabled health care.

(JMIR Hum Factors 2023;10:e45166) doi:10.2196/45166

KEYWORDS
digital health; remote patient monitoring; RPM; human-centered design; human-computer interaction; implementation science
**Introduction**

Networked medical devices offer the potential for people affected by a variety of chronic conditions to monitor symptoms and physiological measures at home, and for the clinicians who treat them to gain more fine-grained and nuanced insight into their lived experience beyond visits to the clinic. Typically, this form of mobile health care (mHealth) is referred to as remote patient monitoring (RPM), defined as “the use of a non-invasive, wearable device that automatically transmits data to a web portal or mobile app for patient self-monitoring and health provider assessment and clinical decision-making” [1]. RPM proponents highlight opportunities for improved patient outcomes, decreased costs, and increased physician satisfaction [2-4]. It is also suggested that RPM will improve the timeliness of care, increase treatment adherence, and support personalized preventive medicine [5-7]. Recently, 2 key drivers have provided a strong motivation for health care practitioners in the United States to adopt RPM options as part of a growth in remote “virtual-first” health care offerings: the approval in 2018 for RPM to be reimbursed through the Centers for Medicare and Medicaid Services, providing the financial support of the largest US health care payer [8], and the rapid shift to remote provision of health care experienced during the COVID-19 pandemic [9-11].

Despite the growing enthusiasm for remote-supported clinical care delivery, to date, large-scale RPM implementation within health systems has been limited. Most studies on the use of RPM technology focus on smaller one-off initiatives—such as grant-funded research studies and disease- or department-specific pilot projects—and much of the evaluation focuses on proving clinical effectiveness in controlled settings and identifying issues with study quality (eg, the inability to conduct double blind trials, and study heterogeneity) [1,12]. As with digital health technology in general, the pragmatic use of RPM in clinical practice has been limited by issues of usability and acceptability, appropriateness for real-world disease management, integration into clinical and technical workflows, and cost-effectiveness [13]. Additionally, few best practices or practical guidelines exist to support the real-world implementation of RPM, or to set a foundation for successful use of this technology at-scale.

In response to this identified gap, health care services researchers have called for better understanding of the health care technology paradigm, including the need to design “person-centered” models that incorporate the needs and experiences of various stakeholders and are built with scalability and sustainability in mind. In this study, we explore the early implementation of a pragmatic operational RPM initiative across the ambulatory network of an academic medical center in New York City, focusing on the experiences, perceptions, and needs of patients and clinicians, with the goal of identifying key “person-centered” themes that can inform implementation recommendations for health systems interested in effectively deploying this technology.

**Methods**

**Study Design**

Our research used a multimethod qualitative approach consisting of semistructured interviews and design thinking workshops among participants (clinical staff, patients, and project implementation team members) of an ambulatory RPM initiative. These activities were conducted as part of a health information technology operational initiative to systematically expand the use of RPM technology to support blood pressure management for hypertensive patients within the expanded ambulatory network of one of the largest academic medical centers in the northeast United States.

**Population and Setting**

The New York University Langone Health (NYULH) system is a large urban academic institution and tertiary care center, with a network of more than 15,000 clinicians in over 400 locations across New York, New Jersey, and Florida. The ambulatory networks consist of academic practices, community-based practices, and federally qualified health centers (FQHCs), as well as ambulatory surgery and rehabilitation centers. NYULH serves an ethnically and socially diverse population with a broad payer mix.

As part of the health system’s overall patient digital experience efforts, NYULH has invested in operational support to facilitate pragmatic implementation of RPM across its practices. NYULH’s electronic health record (EHR), Epic, supports both a “native” RPM integration—allowing data from Bluetooth-connected devices to “stream” through smartphones (eg, Android/GoogleFit or Apple/Healthkit) to the EHR—as well as manual data upload via the MyChart patient portal to support patients without smartphone access. MyChart also serves as a patient-facing tool to integrate RPM education and task management. RPM initiatives within the NYULH system are supported by the Medical Center Information Technology (MCIT) department, which specializes in the analysis, development, and delivery of applications and enterprise information technology solutions and includes personnel with expertise in software development (back-end and front-end programming), solution architecture, design, quality assurance, infrastructure engineering, and product operations and support (including Epic integration).

**Data Collection**

**Semistructured Interviews With RPM “Early Adopter” Clinicians**

In 1962, communication theorist Everett Rogers introduced the concept of innovation “early adopters,” referring to the percentage of individuals who are quick to adopt a new technology, product, or idea [14]; these individuals offer unique perspectives and are considered integral to an innovation’s larger success. In health care, understanding the experiences of early innovation adopters can help facilitate the translation of these tools into diverse environments and support effective, safe, and sustained use.
In this study, early adopters of RPM technology were clinicians in the metro New York practice network (Manhattan, Brooklyn, and western Long Island) who used RPM between 2018 and 2021, prior to the health system’s system-wide RPM initiative. This included maternal-fetal medicine (blood pressure monitoring), pediatric endocrinology (continuous glucose monitoring), lung transplant (temperature, spirometry, pulse oximetry, and weight), bariatric surgery (weight, blood pressure, and glucometer), and primary care FQHCs (blood pressure), each of which independently piloted monitoring initiatives as part of grant-related quality or clinical effectiveness initiatives with the support of the MCIT team. Clinicians at practices with early experience using RPM technologies in disease management were invited to participate in a 30-minute semistructured interview as part of an RPM operational quality improvement effort to identify early resource needs and potentially successful strategies to inform scaling RPM more broadly within the ambulatory health system (see Multimedia Appendix 1). Interview prompts were guided by literature review and focused on clinician experiences, workflows (clinical and technical), and barriers and facilitators. Prompts were reviewed with content experts, the RPM implementation team, and ambulatory clinical leaders. The interviews were conducted in December 2020 and May 2021. Interviews were audio recorded and detailed notes were also taken.

Speculative Design Workshops

Speculative design and futuring workshops have emerged as popular research practices in human-centered and human-computer interaction design. Building on critical design practices and design fictions [15,16], these workshops aim to spark discussion and encourage reflection on the potential implications of emerging technologies. Workshop activities provide a creative environment in which participants can safely explore potentially challenging topics, without a commitment to developing practical solutions to their real-world problems. Outputs may include visual or textual narratives, or prototypes that illustrate alternate futures.

For this research, four 90-minute speculative design workshops were developed to study the following patient identities: (1) early adopters of health technology; (2) patients with chronic diseases (eg, diabetes); (3) parents; and (4) other caregivers. Patients aged between 30 and 76 years were recruited from New York metropolitan area and New Jersey. The workshops were adapted from the speculative design approach in More&More Unlimited’s Investing in Futures framework, which focuses on envisioning future-facing case scenarios and experiences using RPM technology (see Multimedia Appendix 2) [17]. Each session began with participants sharing care experiences, specifically positive encounters with health care providers and moments when they felt cared for. The second activity was world building, here we asked participants to wonder together about what a good future might look like using cards with prompts. The prompts were divided into 8 categories: wearables, health management, smart homes, community, data storage, communication, cost, and health care system. The third and last activity was to build a day in the life of 1 person in the new future using their assigned patient identity. Sessions were conducted from March to August 2021. Transcripts from the workshops were used to identify themes and collect quotes from participating patients.

Data Analysis

This study followed the Standards for Reporting Qualitative Research reporting guideline for qualitative studies [18]. All data sources were recorded (audio logged and transcribed) and deidentified prior to analysis. For qualitative analysis, we applied a hybrid inductive-deductive approach described by Fereday and Muir-Cochrane [19]. First, the data were analyzed using an inductive coding process to identify and iteratively develop and refine emergent themes and codes. Subsequently, a deductive approach focusing on barriers and facilitators was applied to elucidate particular themes relevant to challenges in early RPM implementation. A representative subset of data (2 interviews and 1 speculative design session) was independently coded by 3 primary coders (LLG, ZJ, and NS). Codes were iteratively discussed with the larger research group to (1) review major and minor themes and points of divergence and convergence, (2) establish and refine the code book, and (3) determine thematic saturation. Data were then independently recoded by the primary coders. An additional coder (GD) independently read the coded data for accuracy and to identify cross-cutting themes.

Ethical Considerations

This study was conducted as part of a quality improvement and patient safety evaluation in conjunction with the NYULH MCIT Department. Researchers completed an NYU Langone Health Institutional Review Board–approved quality improvement self-certification.

Results

Overview

Twenty-four stakeholders participated in the study across data collection methodologies. Thirteen early adopter clinicians (n=11 physicians, n=1 nurse practitioner, n=1 registered nurse; n=7, 53% identified as female; average 16 years in clinical practice) represented pediatric endocrinology, maternal-fetal medicine, weight management clinics, pulmonary transplant, internal medicine, and federally qualified health centers. Twenty-one patient representatives participated in 4 speculative design sessions (n=5 early adopters, n=5 parents, n=5 caregivers, and n=6 managing a chronic condition; n=11, 53% identified as female).

We present our findings across three main themes: (1) data collection and practices; (2) proactive and preventive care; and (3) health disparities and equity. We also identify evidence for mitigation and support to address challenges in these areas (Textbox 1 and Multimedia Appendix 3).
Textbox 1. Themes and subthemes identified.

<table>
<thead>
<tr>
<th>Data collection and practice</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Clinical impacts of real-world measures</td>
</tr>
<tr>
<td>• Issues of data sharing, security, and privacy</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Proactive and preventive care</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Proactive and preventive monitoring</td>
</tr>
<tr>
<td>• Proactive interventions and support</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Health disparities and equity</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Tailored and flexible care</td>
</tr>
<tr>
<td>• Implicit bias</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mitigation and support for remote patient monitoring-enabled health care</th>
</tr>
</thead>
</table>

Data Collection and Practice

Overview

The first finding centered on experiences and concerns regarding sharing data about personal health measures that are generated in settings other than the clinic. Two areas in particular were highlighted among participants: (1) clinical impacts of real-world measures and (2) issues of data sharing, security, and privacy.

Clinical Impacts of Real-World Measures

The potential for RPM to expand clinical diagnosis and management capabilities through the collection of more ecologically valid measures in a wider range of real-world contexts that better reflect a patient’s lived experience were noted by both clinicians and patients.

My mom gets really anxious around doctors, and being in a doctor’s office, and so it always looks like her blood pressure is through the roof when she’s there. She has to manually track her blood pressure at other times so she can go to the doctor and be like, “No I’m not. This isn’t my standard. I have a very normal blood pressure. I’m just a little freaked out by you trying to give me blood pressure medication.”

[Speculative design session 1, participant #2, female]

A physician noted similar data quality worries, pointing in particular to concerns about older patients being prescribed multiple medications for hypertension and the challenge of using clinical blood pressure values to accurately gauge the overall effects of the medication:

Allowing a patient to log at home, particularly for older patients... Some patients read high in clinic and are fine at home. [Clinician #6, male]

At the same time, participants also noted concerns regarding the feasibility of using remote monitoring technologies, and the burdens placed on patients to routinely collect these data, particularly around access, language, and digital and health literacies.

Tech access and literacy is a global concern.... [Patients] may not be able to use RPM technologies to accurately self-measure blood pressure or glucose.

[Clinician #6, male]

This impacted the clinician’s confidence in using the patient-reported data to make clinical decisions. Additionally, attending to RPM health data can have unique psychological impacts on patients, both positive and negative; on one hand, patients responded positively to data within “normal” values, as it provided a reassurance that everything is going well, however, when the data suggested something outside of expectations it was seen as a source of worry for patients:

I think it is a little bit of some mental warfare for [patients], because if their number is a little low and they can’t get the number they know, then they’re obviously worried that there’s some problem...they’re worried about [transplant] rejection. [Clinician #2, female]

The physician noted being unsure of the best way to counsel patients on the impacts of this “mental warfare,” or how to adjust home monitoring to optimize patient well-being.

Issues of Data Sharing, Security, and Privacy

A subtheme of particular importance to patient participants was that of the relationship between RPM and data sharing, security, and privacy. For many patient participants, RPM data represented an opportunity for better connected health care experience. In particular, there was interest in the capacity to share data between primary care physicians and medical specialists:

You know how when you get older you see one specialist after another? Specialists are starting to be able to coordinate MyChart [the patient portal] and all these other things. But a lot of doctors, well you know, I uploaded it to this one and that one, I ran through permission but then by the time you get to the doctor’s office it’s like “I can’t see your records.” [Speculative design session 3, participant #3, female]

For other patients, there is an apparent trade-off between the benefits of data connectedness and the risks of privacy breaches:
I know that there are significant privacy issues with these [digital health tools], but I feel like it's an area, especially with all of these wearable devices and everything, it just makes sense to begin to connect more, to be able to pull it together and get a better level of care as a result. [Speculative design session 4, participant #2, male]

For others, data collection and sharing raised significant questions about privacy and security, in particular regarding how data might be used by or shared with other companies:

I just feel like companies that will be collecting all this information, let's say in a future scenario monetary system, what if they're selling your data to third parties? That would really kind of be a concern. [Speculative design session 2, participant #4, female]

As a result of these impressions, many patients expressed reservations about using RPM technologies regularly in their care. Conversely, clinicians did not routinely mention data privacy or security as an issue in their data management or clinical practices, focusing instead on the aforementioned challenges of data quality and interpretability in clinical contexts.

Proactive and Preventive Care

Overview

A specific area of reflection from both patients and providers revolved around the impact of RPM on the practice of providing medical care, and the potential shift from reactive to proactive care provision enabled by the technology. Subthemes on this topic include: (1) proactive and preventive monitoring and (2) proactive interventions.

Proactive and Preventive Monitoring

Clinicians discussed a number of potential opportunities that RPM data might facilitate with regard to population health and preventive monitoring of their patients, such as the ability to programmatically identify patients that are struggling to maintain a suggested program of treatment or who are not being adequately served. In this way, they would hope to reduce the likelihood that these patients would end up in the emergency room (clinician #6, male). Similarly, the potential for clinicians to proactively communicate with their patients in response to data generated by RPM-enabled health care was highlighted as a key benefit by multiple providers. One physician pointed to the way that blood pressure data come directly as a message to their EHR in basket, allowing them to respond using the patient portal (clinician #1, female). When a series of readings were considered beyond normal values, clinicians commented they could reach out immediately to the patient and ask them to schedule a visit more promptly than they would have otherwise (clinician #7, male). At the same time, clinicians also drew attention to concerns that patients who do not actively submit their remote health data might receive less proactive attention from clinicians, and potentially be excluded from outreach initiatives that relied on these data to identify patients. For example, if a patient fails to submit remote blood pressure data in a particular week, they will not receive a call from the nurse, as in this workflow they would not be identified in the EHR (clinician #7, male). Clinicians also noted the unintended consequences of the increase in patient-generated messaging around their RPM data: “It can take several hours to go through everything [in the EHR]” (clinician #1, female).

From the patient perspective, a key challenge highlighted during the workshops was the potential negative effect of this proactive communication:

Maybe somebody doesn’t want the doctor calling them every time there’s a little spike... They’ll be like, everything’s fine, leave me alone, I’ll call you if there’s a problem. I can kind of see that maybe being a little invasive. [Speculative design session 2, participant #4, female]

There was broad concern among participants in the patient workshops that an increased reliance on data and technology to monitor progress might lead to a reduction in personal care and remove opportunities for empathetic connections: “My biggest concern with this would be that if doctors are getting all these numbers all the time, it doesn’t dehumanize you” (speculative design session 1, participant #4, male):

You know, once we start adding more machines and more technology, people lose that personal sense of connection and that’s just something I’m not willing to sacrifice. [Speculative design session 2, participant #3, female]

Proactive Interventions and Support

Building on the opportunities offered by proactive communication, providers also discussed how RPM can support proactive interventions, specifically between scheduled visits:

[Before RPM] if patients were just getting started on medications the [staff] would bring them back next week to review logs.... Now they don’t have to do that, they can still see them in two weeks, and can [review remotely] in between visits. [Clinician #2, female]

For 1 clinician, RPM promised opportunities to specifically support health behavior education:

I’m hoping it helps in the sense that we’re really able to get our patients involved in their care, and it’s not just every three months I get my sugar checked when I come to the clinic. If they’re doing it on a daily basis then it helps them to realize it’s important and the education helps them understand that everything they do impacts their health. I’m hoping that it helps our patients to understand that everything they do really does make a difference.... I’m hoping it helps us track and make patients aware of their choices. [Clinician #4, female]

Patients were keen to highlight how data from RPM might also help in their own decision-making to support healthier choices and behavior:

And so the more data that I have I feel like I can make better decisions, right? Something that I was thinking about doing was making a food journal, because I really don’t know how many calories I’m actually...
Another patient postulated that RPM data might provide the basis for an insurance incentive to encourage people to follow through with suggestions made during their annual check-up (speculative design session 1, participant #5, male).

While proactive interventions were generally considered a positive opportunity, providers did highlight challenges in this new workflow, including billing and reimbursement. This was particularly relevant among the FQHC providers:

“FQHCs are a little different, we can’t bill directly for RPM but we want to use the technology,... I’m concerned about sustainability. If we can’t bill for it, how does it pay for itself?” [Clinician #5, male]

Experiences in other practices led clinicians to comment on the tendency of patients to disregard RPM data, for example by switching data streams off rather than positively responding and adopting healthier behaviors, thereby limiting its potential effectiveness (clinician #1, female). Overall, clinicians expressed concerns about being asked to provide more care as a result of RPM, “We’re asking them to do more work between visits when they’re not compensated, and that’s hard” (clinician #5, male).

Health Disparities and Equity

Overview

Both clinicians and patients discussed the role of RPM technologies in addressing or exacerbating inequalities in health care, with the main areas of interest and concern regarding (1) tailored and flexible care and (2) implicit bias.

Tailored and Flexible Care

For clinicians, RPM presents an opportunity to directly address health inequities through expanded access to care and more effective, tailored health care usage. One clinician discussed the potential of targeting care delivery to specific underserved women of color who were disproportionately affected by gestational diabetes and could benefit from tailored monitoring (clinician #2, female). The opportunity to replace arduous visits to the clinic with a telehealth virtual visit was considered an incentive for patients to present to and stay in care:

“For patients [living] in other boroughs it’s so hard to come into Manhattan. Parking. If their family member takes them. It’s a lot. So, we will sometimes do video visits for people in this area who just [can’t] come in. [clinician #8, female]

In the context of RPM specifically:

“We kind of, for better or worse, use [RPM] a little bit like a reward system.... If you’re not doing your monitoring you have to come in, because we can’t do a proper visit with you.” [clinician #8, female]

However, this potential of RPM to address health disparities was viewed as potentially limited for a number of reasons, including barriers related to the technology and its appropriateness or easy use for diverse patients. A lack of culturally and contextually congruent technology, and wrap around services for RPM-enabled health care, was considered a major barrier to its ability to be effectively tailored to diverse patients who may otherwise have benefited:

“My main concern is that it’s not all in Spanish. The majority of our patients speak Spanish. [Clinician #4, female]

Other participants reflected that, in order to support flexible and tailored care for diverse patients, patients and clinicians need to overcome a range of barriers already associated with health technologies. For example, many of the RPM devices that integrate most effectively with the hospital’s EHR system are only available through more expensive and comprehensive insurance plans, and so may be unavailable to patients that could benefit the most (clinician #3, female). Similarly, 1 physician described how current billing mechanisms for RPM-enabled health care often result in out of pocket fees, which make it an inaccessible option for those with limited income (clinician #1, female).

Implicit Bias

A particular area of challenge in RPM-enabled health care practice is centered around identifying patients who might be considered good candidates for RPM. Clinicians noted that they thought certain patients were probably better suited for RPM programs than others, often based on individual assessments of digital literacy skills, health literacy, language abilities, and proactive participation in care. This was highlighted by a physician who said that patients would most likely be selected for RPM programs based on the question, “Do I think they can do it?” explaining that this would be a soft assessment that takes into consideration financial concerns as well as digital literacy and health literacy (clinician #7, male). For another physician, patients considered a potentially bad fit for RPM would be “people who don’t want to use technology” or who faced “language barriers” (clinician #2, female). A third clinician reported:

“Tech-savvy persons and high-literacy people are more likely to use [RPM]...[we are] less likely to offer RPM to people who are less likely to use it.” [Clinician #4, female]

This process of patient identification and selection for inclusion in RPM programs was not explicitly identified to be a form of systematic bias by either clinicians or patients; rather, it was most often discussed as a factor that contributed to the patient’s activity or success within the RPM programs.

Mitigation and Support for RPM-Enabled Health Care

Participants identified possible approaches to supporting opportunities and mitigating challenges posed by novel RPM practices. These typically related to new roles and responsibilities that might be created to better support interactions with technology systems and health care administration. For example, 1 patient highlighted how cultural competence and concordance can be important to equity in health care delivery:

“I have over the past few years intentionally sought out doctors who were people of color, particularly Black Americans and doctors who were women. I just
find that in doing so, I think my health care in general is usually better. There’s questions and discussions and cultural sensitivities that I find are being addressed in general when I have doctors of color and doctors who are women. [Speculative design session 4, participant #5, female]

Clinicians noted that having a range of clinical staff who speak a patient’s first language may be critical to RPM, as it demands a higher frequency of communication and support (clinician #4, female).

Both clinical providers and patients highlighted a variety of roles or services that could be provided to support patient navigation, advocacy, and competence. One patient stated:

I feel like one of the things about the health care system right now is that it is so confusing to read about your benefits or your insurance and what’s covered by what and how you qualify for things. [Speculative design session 1, participant #2, female]

Clinicians cited community health workers (CHWs) as helpful to connecting with vulnerable populations, in part because they may already be making home visits with these patients (clinician #4, female). In particular, CHWs were considered potential digital advocates who could act as intermediaries with RPM technology:

CHWs [could perform] teaching around what to expect and how to use the [devices], assessing access to wireless communications...to help them troubleshoot. [Clinician #6, male]

However, it was noted that additional training would be required for this staff, as these tasks were considered outside of the current scope of their work:

CHWs are great to work with. They can help facilitate the MyChart sign up and encourage them to use [RPM] in a way that is helpful. It would definitely require some training in how to use it, though. [Clinician #5, male]

**Discussion**

**Summary of Results**

In this study, we present the challenges and considerations associated with the transition of a health care system to a care delivery model enabled by RPM technology. Using data triangulated from an institution currently undergoing the pragmatic deployment and scaling of RPM in practice, we identified 3 main themes of interest and concern to RPM stakeholders: (1) novel data collection practices and concerns; (2) proactive and preventive care models; and (3) health disparities and health equity. Our work also identified opportunities for mitigation and support for RPM-enabled care, particularly around new work roles and resources to support those who are engaging in this type of care delivery. This study contributes to the existing literature on remote monitoring by capturing the experiences and perspectives of various key stakeholders (including clinicians, patients, caregivers, other health staff) in a health care system that is actively undergoing a care delivery transition enabled by RPM technology.

**Implications for the Pragmatic Deployment of RPM Within Health Systems**

This work highlights a few key areas of consideration for health care systems or practices that are considering undertaking RPM-enabled care transformations. The first is around ensuring a “successful” remote monitoring experience for patients and clinicians. Substantial research in the areas of digital health technology points to the clear need for an improved overall user experience of these technologies, from back-end data integration and interoperability to front-end product design [20-22]. As endorsed by both patients and clinicians in this study, the ideal RPM experience is driven by the efficient transmission of validated, trustworthy data in an environment that is user-friendly, humanistic, and information secure. At the same time, priorities of different stakeholders regarding these technologies may not overlap, and in some instances may be in direct conflict. Our study showed that, while clinicians and patients discussed similar themes regarding RPM, their individual concerns or perspectives were often conflicting—while clinicians endorsed wanting to have more ability to access data for patient care, patients were concerned about their clinicians monitoring them continuously; and while clinicians expressed concerns about being overly contacted by their patients between visits, patients themselves were worried about losing their personal relationships with their doctors. Clearly identifying and, where possible, aligning the diverse needs and preferences of RPM stakeholders can facilitate a more acceptable, usable RPM program; it can also reduce potential areas of friction that might contribute to nonadoption, abandonment, or unintended negative effects on patient experience or patient-clinician relationships.

The second area highlights the novel practice structures that can be implemented to better support RPM-enabled care; specifically, the opportunity for new work roles around digital health navigation and advocacy, and the shifted nature of the patient-clinician relationship. In our study, both patients and clinicians identified the potential that digital advocates (including CHWs) might have on improving the experience of care using RPM. Our findings reflect other work indicating how the introduction of data-intensive technology such as RPM can bring with it important changes to clinicians’ work, including increasing administrative labor and shifting temporal work patterns (eg, work outside work or “pajama time”) [23-25]. New practices that empower members of the health care team beyond physicians, or that automate the more routine aspects of these interactions while allowing physicians to engage in specialized practice, will be needed to successfully implement data- or monitoring-intensive technologies [26]. Examples that have been suggested elsewhere include the use of AI-enabled chatbots to help perform clinical analyses and provide an initial response to data that are within expected ranges, and the implementation of big data analytics and machine learning more widely within health care information technology [27]. However, it should also be noted that participants in the patient workshops strongly indicated how important a close relationship with their clinical provider was for them; patients would prefer to securely share
data with the clinicians they know and trust and make collaborative decisions with empathetic physicians. Similarly, while clinicians are keen to encourage patients’ engagement in their own care, they too indicated how acting on RPM data should be collaborative. These findings echo others in the digital health literature as well as the larger human-computer interaction research field, which warns of the danger of focusing too strongly on data and highlights the positive impacts of more emotional and experiential self-reporting to health and well-being [28-31]. This suggests that, at its core, technology-centered care such as RPM should include an element of humanistic, mutually beneficial comanagement, which should include not only the patient-provider-technology triad, but the expanded team of digital care advocates (both trained and lay) as well.

The third area of critical consideration is that of “inclusive RPM.” Integrating health equity considerations into interventions and prevention programs has been identified by the US Centers for Disease Control and Prevention as a key factor in improving public health [32]. While digital health technologies such as RPM have the potential to improve health equity, our findings also reflect concerns that the uneven application of these technologies in clinical contexts may exacerbate existing health disparities or potentially create new sources of digitally-mediated inequities [33]. Our findings also indicate that the promise RPM technologies offer toward directly addressing health inequities by expanding access and tailoring health care may be more fragile outside the constraints of research studies. Our findings highlight clinicians’ expectations for RPM technologies to help facilitate a more flexible and tailored approach to health care provision, which includes virtual appointments and interventions, and which reduces the impacts of travel and unpaid leave that disproportionately impact patients from underserved communities. However, we also identified a number of barriers to implementing RPM-enabled care in a way that effectively addresses disparities. These barriers are consistent with existing literature showing that digital health care interventions are generally more accessible to socioeconomically advantaged groups, and that health technology programs often neglect digitally-mediated factors at the community or society level which influence health disparities [34,35].

Limitations
This study has several limitations. It reflects the experience of a particular health care institution at a given moment in the pragmatic development of its remote monitoring program. Participants were identified through pragmatic convenience sampling methodologies and may not reflect the overall composition of either the institution itself or the larger pool of patients and providers engaging with RPM. Data were collected through a variety of methods and may reflect a number of biases, including interviewer bias and response bias. Due to small sample sizes and risks of participant identification, we are unable to provide more detailed information on specific clinic- or department-level experiences that may have differed between practices. Strengths of the study are that it reflects the experiences of key stakeholders (patients and clinical providers) participating in the real-world implementation of a digital health intervention, using an organic multisource qualitative approach to ensure a diversity of stakeholders, approaches, and contexts were captured.

Conclusions
In this paper, we present an inquiry into the challenges and considerations associated with the transition of RPM-enabled health care from research studies into clinical practice. Our analysis of qualitative data from patients, clinicians, and health staff identified 3 main themes related to the pragmatic implementation of this technology, including issues around data collection and review practices, proactive and preventive care experiences, and technology-mediated health disparities and inequity. We further identified opportunities for mitigation and support of the challenges and opportunities raised, including building skills, capacity, and diversity among the future clinical workforce engaged in RPM-related care. Ultimately, the introduction of RPM-enabled health care poses particular design and implementation challenges for current practices, creating a potentially unbalanced patient-provider-technology triad that can disrupt practice patterns and norms and affect the experience of care for both patients and clinicians. Understanding and responding to these challenges can help improve its acceptability, scaled use, and sustainability in health care delivery.
Multimedia Appendix 3
Themes, subthemes, and representative quotes.

References


Abbreviations

CHW: community health worker
EHR: electronic health record
FQHC: federally qualified health center
MCIT: Medical Center Information Technology
NYULH: New York University Langone Health
RPM: remote patient monitoring
Accuracy and Reliability of a Suite of Digital Measures of Walking Generated Using a Wrist-Worn Sensor in Healthy Individuals: Performance Characterization Study

Nathan Kowahl1*, MS; Sooyoon Shin1*, PhD; Poulami Barman1, MS; Erin Rainaldi1, MS; Sara Popham1, PhD; Ritu Kapur1, PhD
Verily Life Sciences, South San Francisco, CA, United States
*these authors contributed equally

Corresponding Author:
Nathan Kowahl, MS
Verily Life Sciences
269 E Grand Ave
South San Francisco, CA, 94080
United States
Phone: 1 415 504 2681
Email: natekowahl@verily.com

Abstract

Background: Mobility is a meaningful aspect of an individual’s health whose quantification can provide clinical insights. Wearable sensor technology can quantify walking behaviors (a key aspect of mobility) through continuous passive monitoring.

Objective: Our objective was to characterize the analytical performance (accuracy and reliability) of a suite of digital measures of walking behaviors as critical aspects in the practical implementation of digital measures into clinical studies.

Methods: We collected data from a wrist-worn device (the Verily Study Watch) worn for multiple days by a cohort of volunteer participants without a history of gait or walking impairment in a real-world setting. On the basis of step measurements computed in 10-second epochs from sensor data, we generated individual daily aggregates (participant-days) to derive a suite of measures of walking: step count, walking bout duration, number of total walking bouts, number of long walking bouts, number of short walking bouts, peak 30-minute walking cadence, and peak 30-minute walking pace. To characterize the accuracy of the measures, we examined agreement with truth labels generated by a concurrent, ankle-worn, reference device (Modus StepWatch 4) with known low error, calculating the following metrics: intraclass correlation coefficient (ICC), Pearson r coefficient, mean error, and mean absolute error. To characterize the reliability, we developed a novel approach to identify the time to reach a reliable readout (time to reliability) for each measure. This was accomplished by computing mean values over aggregation scopes ranging from 1 to 30 days and analyzing test-retest reliability based on ICCs between adjacent (nonoverlapping) time windows for each measure.

Results: In the accuracy characterization, we collected data for a total of 162 participant-days from a testing cohort (n=35 participants; median observation time 5 days). Agreement with the reference device–based readouts in the testing subcohort (n=35) for the 8 measurements under evaluation, as reflected by ICCs, ranged between 0.7 and 0.9; Pearson r values were all greater than 0.75, and all reached statistical significance (P<.001). For the time-to-reliability characterization, we collected data for a total of 15,120 participant-days (overall cohort N=234; median observation time 119 days). All digital measures achieved an ICC between adjacent readouts of >0.75 by 16 days of wear time.

Conclusions: We characterized the accuracy and reliability of a suite of digital measures that provides comprehensive information about walking behaviors in real-world settings. These results, which report the level of agreement with high-accuracy reference labels and the time duration required to establish reliable measure readouts, can guide the practical implementation of these measures into clinical studies. Well-characterized tools to quantify walking behaviors in research contexts can provide valuable clinical information about general population cohorts and patients with specific conditions.

JMIR Hum Factors 2023;10:e48270 doi:10.2196/48270
KEYWORDS
digital measurements; wearable technology; mobility measurements; walking patterns; wearable; wearables; sensor; sensors; mobility; measurement; measurements; walk; walking; gait; step; wrist-worn; reliability; accuracy

Introduction
Assessing an individual’s mobility can provide meaningful insights into their general health status. In clinical settings, mobility is a fundamental factor to define prognosis and care as it is closely associated with a wide array of health outcomes [1-3]. However, accurate and reliable quantification of mobility in real-world settings remains challenging because self-reported data from instruments such as the International Physical Activity Questionnaire can be biased by limited recall and social desirability [4,5].

The interest in quantifying physical activity using wearable devices has recently increased, as these technologies can collect objective individualized data [6]. Wearable sensors have been incorporated into clinical studies across different disease states to enable movement analyses and the quantification of discrete physical activities to develop clinically meaningful end points [7,8].

Yet, to cement their research utility, two aspects of these digital measurements need to be properly characterized: (1) the accuracy with which a digital measurement reads the parameters of interest [9] and (2) the amount of aggregated data needed to reliably capture an individual’s underlying behavioral state, minimizing noise related to natural variability, which usually translates into an aggregation time period for data collection (time to reliability). Although accuracy is always a critical aspect in the characterization of a measure’s performance, time to reliability tends to be ignored, even though it is key for establishing fundamental study design specifications (eg, collection time periods and length of wear time per day), defining baselines, or computing power calculations for the detection of intervention effects or other changes.

Studies characterizing the performance of digital walking measures often focus on step count. However, the literature around these studies shows considerable heterogeneity across designs and some notable limitations. First, analyses tend to rely on truth labels originated by participants’ self-reports, short-term close monitoring [10-12], or from reference devices with suboptimal accuracy (mean absolute percentage error >20%) and with the same body placement as the investigational devices, which would bias agreement results [13]. Second, these studies are often conducted in artificial laboratory environments, which inherently limit behavior range and are susceptible to subjectivity, assessment bias, and unreliability [14-16]. Third, reliability characterization for investigational digital measurements is often absent from studies, despite having been acknowledged as an important element for the validation of clinically important research metrics, such as patient-reported outcomes [17]. Beyond step counts, there have been studies that have used other digital measures (eg, walking intensity captured by the peak 30-minute cadence) to generate clinical insights but without full characterization of their performance [18-29].

In a previous study, we developed an algorithm that accurately classifies ambulatory status from data collected from a wrist-worn device, characterizing its performance across diverse demographic groups in a real-world setting [30]. Further, results from a substudy of an interventional randomized phase 2 trial demonstrated that digital measures of physical activity (step count and ambulatory time) could be sensitive to treatment effects in patients with Lewy body dementia [31].

Herein, we report on the development of a series of measures that can capture walking behavior comprehensively, characterizing their analytical performance in accuracy and reliability. These measures included (1) step count, (2) walking bout duration, (3) number of total walking bouts, (4) number of long walking bouts, (5) number of short walking bouts, (6) peak 30-minute walking cadence, and (7) peak 30-minute walking pace. To characterize their accuracy, we compared the measure readouts generated from a study device with highly accurate truth labels from an ankle-worn reference device in healthy volunteers. To characterize their reliability, we developed a novel approach to calculate the aggregated time required to reach a reliable readout (time to reliability) for each measure.

Methods

Study Participants
The study cohort (pilot program study) included adult volunteer participants, recruited among Verily Life Sciences employees in 2 locations (South San Francisco, CA, and Cambridge, MA), without specific selection criteria. Gender and age information was collected for the accuracy characterization (not for the reliability characterization). This study was determined to be exempt research that did not require institutional review board review.

Devices
The Verily Study Watch was the study device. This is a wrist-worn smartwatch that records acceleration data via an onboard inertial measurement unit with a 30 Hz 3-axis accelerometer. The study device also has a photoplethysmography sensor and an additional accelerometer and gyroscope with a 100 to 200 Hz sample rate, which this study did not use.

For the accuracy characterization, we used the ankle-worn Modus StepWatch 4, a Food and Drug Administration–listed, 200 Hz 3-axis accelerometer device, as a reference to obtain ground truth labels for step counts (and, subsequently, the other derived walking measures); raw acceleration data from this device were not used for this study. This device has shown the greatest accuracy for step counting relative to other wearable devices compared with human counting in real-world and in-lab settings [24,32].
Generation of Digital Measurements

We collected continuous, raw accelerometer sensor data from the study smartwatch, computing step counts for every 10-second, nonoverlapping epoch (for additional information about the algorithm associated with the study device to determine step counts, see Multimedia Appendix 1), and collected step count outputs from the reference device (generated by the algorithm associated with the StepWatch) also in 10-second epochs. From the 10-second epoch-based step counts, other measures of walking were derived, applying the same computations to the step counts from both devices (summarized in Table 1). We report the measure readouts as daily aggregates for individual participants (ie, participant-days).

Table 1. Summary of walking measure definitions.

<table>
<thead>
<tr>
<th>Daily walking measure</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step count</td>
<td>Summed number of steps per day</td>
</tr>
<tr>
<td>Number of walking bouts&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Summed number of walking bouts per participant-day</td>
</tr>
<tr>
<td>Number of short walking bouts</td>
<td>Summed number of walking bouts lasting between ≥20 seconds and &lt;1 minute, per participant-day</td>
</tr>
<tr>
<td>Number of long walking bouts</td>
<td>Summed number of walking bouts lasting ≥2 minutes per participant-day</td>
</tr>
<tr>
<td>Walking bout duration, mean</td>
<td>Mean duration of daily walking bouts</td>
</tr>
<tr>
<td>Walking bout duration, SD</td>
<td>SD of the duration of daily walking bouts</td>
</tr>
<tr>
<td>Walking bout duration, 95th percentile</td>
<td>Highest bout duration below the top 5% longest bouts</td>
</tr>
<tr>
<td>Peak 30-minute walking cadence&lt;sup&gt;b&lt;/sup&gt;</td>
<td>For each participant-day, average cadence for the 180 ten-second epochs (ie, 30 minutes, not necessarily contiguous) with the highest cadence</td>
</tr>
<tr>
<td>Peak 30-minute walking pace</td>
<td>For each participant-day, average pace (calculated from the cadence and estimated stride length based on gender and height) for the 180 ten-second epochs (ie, 30 minutes) with the highest cadence (namely, the daily peak 30-minute cadence); measured as meter/second</td>
</tr>
</tbody>
</table>

<sup>a</sup>A walking bout was defined as a series of contiguous 10-second epochs containing ≥2 steps each and lasting for ≥30 seconds (ie, at least 3 epochs). Epochs were considered contiguous if they were not interrupted by >20 seconds (ie, by no more than two 10-second epochs).

<sup>b</sup>Walking cadence was defined as the number of steps per unit of time (in this study, per second—steps/second).

Analyses

Accuracy Characterization

For the characterization of accuracy, the observation period ranged from June 2019 to December 2019. The overall analysis cohort (N=70) was split into 2 equal subcohorts (n=35): training and testing groups (Figure S1 in Multimedia Appendix 1). Participants were required to wear the 2 devices, the smartwatch and the reference device, throughout waking hours for up to 10 days. Step counts were obtained for both the study and the reference devices, for as long as both devices had been worn simultaneously by each participant, and filtered for days with ≥8 hours of wear time and >100 steps. Each subsequent measure was derived based on step counts from each device (Table 1) and compared for agreement. Agreement was examined using the following metrics: Fisher intraclass correlation coefficient (ICC) as the main metric, Pearson r coefficient, mean error, and mean absolute error. For each metric, we calculated 95% CIs by bootstrapping with 1000 resampling iterations to account for multiple days (generally 5) from a given participant. Additionally, to further characterize the degree of agreement and bias of each measure, we examined measurements and distributions between devices and Bland-Altman plots with 95% limits of agreement.

Reliability Characterization

For time-to-reliability characterization, the observation period was 20 months (April 13, 2018, to December 31, 2019). This analysis was designed to determine the duration of time (from 1 to 30 days) over which each measure needs to be aggregated (the different lengths of time tested were termed “aggregation scopes”) to yield stable values, indicating that it reliably captures an individual’s underlying behavioral state. Data were considered analyzable for this objective when participants had worn the device for at least double the duration of a given aggregation scope (in order to have data for 2 nonoverlapping time windows), starting from a minimum of 2 days (for the shortest aggregation scope of 1 day) to a minimum of 60 days (for the longest aggregation scope of 30 days); in addition, at least 50% of the days in each time window had to have ≥12 hours of daily wear. The number of participants meeting these criteria varied according to the span of the aggregation scopes (N=234 for the 1-day aggregation scope [ie, the smallest aggregation scope had the largest cohort]; n=81 for the 30-day aggregation scope [smallest cohort for the largest aggregation scope]; Figure S1B in Multimedia Appendix 1).

In this analysis, we included the same set of measures as for the accuracy characterization, except the 30-minute peak walking pace, because the measure is derived directly from 30-minute peak walking cadence (Table 1); therefore, the results of this analysis were expected to be identical between these 2 measures. We calculated Fisher ICCs between adjacent, nonoverlapping windows of time for each aggregation scope (1-30 days). We computed a rolling mean for each daily aggregated measure over the set number of days for each aggregation scope and then computed the ICC between adjacent windows. We repeated this computation by shifting the start date of each window by 1 day and repeated the computation testing aggregation scopes between 1 and 30 days.
Results

Accuracy Characterization

A total of 162 participant-days worth of data were collected from the 35 participants in the test cohort, with each participant contributing 1 to 10 days (median 5 days). The mean daily step count, daily ambulatory time, and wear time per participant-day were 10,075.88 (SD 4321.07) steps, 1.86 (SD 0.78) hours, and 13.73 (SD 3.00) hours, respectively (Figure S2 in Multimedia Appendix 1).

For each measure of interest (see the Methods section), the comparison of the values generated from the study device against the reference device showed ICC values ranging between 0.701 and 0.865 (Table 2, Figure 1); the measure “mean duration of daily bouts” produced the lowest ICC value (0.701), and “daily step count” had the highest (0.865). Pearson r values were all greater than 0.75, and all values were statistically significant (P<.001; Table 2). The Bland-Altman analysis (Figure 2, middle) revealed that measure differences between the study and reference devices were not dependent on the measure value without significant bias. Scatter plots (Figure 2, left) and distribution (Figure 2, right) of measures between study and reference devices showed overlap for all 9 measures of walking.

Table 2. Summary of results from the characterization of accuracy of the measures generated from the study device compared with those collected from the reference device (N=35).

<table>
<thead>
<tr>
<th>Accuracy metric</th>
<th>ICC(^a) (95% CI)</th>
<th>Pearson r (95% CI)</th>
<th>ME(^b) (95% CI)</th>
<th>MAE(^c) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daily step count</td>
<td>0.865 (0.809 to 0.933)</td>
<td>0.881 (0.832 to 0.941)</td>
<td>151.450 (–486.869 to 726.201)</td>
<td>1643.145 (1196.832 to 1996.216)</td>
</tr>
<tr>
<td>Mean duration of daily bouts</td>
<td>0.701 (0.529 to 0.876)</td>
<td>0.784 (0.657 to 0.922)</td>
<td>14.143 (9.657 to 17.607)</td>
<td>17.141 (12.985 to 20.371)</td>
</tr>
<tr>
<td>Daily bout duration, SD</td>
<td>0.738 (0.525 to 0.918)</td>
<td>0.813 (0.659 to 0.948)</td>
<td>27.813 (17.180 to 36.224)</td>
<td>32.585 (22.038 to 41.081)</td>
</tr>
<tr>
<td>95th percentile of daily bout duration</td>
<td>0.715 (0.433 to 0.918)</td>
<td>0.763 (0.550 to 0.932)</td>
<td>50.293 (28.720 to 67.285)</td>
<td>69.620 (49.051 to 85.290)</td>
</tr>
<tr>
<td>Number of daily bouts</td>
<td>0.756 (0.620 to 0.838)</td>
<td>0.757 (0.632 to 0.848)</td>
<td>-0.611 (–3.740 to 2.840)</td>
<td>11.846 (9.743 to 13.698)</td>
</tr>
<tr>
<td>Number of long daily bouts</td>
<td>0.755 (0.638 to 0.845)</td>
<td>0.781 (0.671 to 0.861)</td>
<td>1.438 (0.780 to 2.103)</td>
<td>2.858 (2.426 to 3.232)</td>
</tr>
<tr>
<td>Number of short daily bouts</td>
<td>0.754 (0.621 to 0.811)</td>
<td>0.768 (0.648 to 0.830)</td>
<td>-2.747 (–4.553 to -0.736)</td>
<td>8.401 (7.172 to 9.542)</td>
</tr>
<tr>
<td>Daily peak 30-minute cadence</td>
<td>0.734 (0.603 to 0.841)</td>
<td>0.773 (0.662 to 0.862)</td>
<td>0.045 (0.013 to 0.076)</td>
<td>0.100 (0.080 to 0.118)</td>
</tr>
<tr>
<td>Daily peak 30-minute pace</td>
<td>0.784 (0.668 to 0.873)</td>
<td>0.802 (0.710 to 0.877)</td>
<td>0.030 (0.007 to 0.051)</td>
<td>0.070 (0.057 to 0.082)</td>
</tr>
</tbody>
</table>

\(^a\)ICC: intraclass correlation coefficient.  
\(^b\)ME: mean error.  
\(^c\)MAE: mean absolute error.
Figure 1. Accuracy characterization: ICC results (and 95% CIs) obtained from the comparison of the digital measurements generated from the study device against those from the reference device. ICC: intraclass correlation coefficient.
Figure 2. Accuracy characterization: detailed results of the comparisons of the digital measures generated from the study device against those from the reference device. Left column: plots of study device readouts (y-axis) versus reference device readouts (x-axis). Middle column: modified Bland-Altman plot showing the difference in mean values between devices (y-axis) versus mean values from the reference device (x-axis). Right column: readout value distributions for both devices in the testing subcohort. (A) Daily step count. (B) Daily walking bout duration, mean. (C) Daily walking bout duration, SD. (D) Daily walking bout duration, 95th percentile. (E) Number of daily walking bouts. (F) Number of daily long walking bouts. (G) Number of daily short walking bouts. (H) Daily peak 30-minute walking cadence. (I) Daily peak 30-minute walking pace. LoA: limits of agreement. See Multimedia Appendix 2 for higher resolution image.

Reliability Characterization

In the cohort of eligible participants who yielded analyzable data (see the Methods section, N=234), individual participant data were collected for up to 596 (median 119) days for a total of 15,120 participant-days (see Figure S1B in Multimedia Appendix 1). The mean daily step count, daily ambulatory time, and daily wear time per participant-day were 9701.06 (SD 4321.88) steps, 77.42 (SD 39.77) minutes, and 17.36 (SD 4.04) hours, respectively.

We defined aggregation scopes of increasing duration from 1 day up to 30 days. For each of these scopes, the participant subcohorts that generated data deemed analyzable were of variable size (generally decreasing as the aggregation scope grew, Figure 3).
Across all the measures of interest in this analysis, the stability of the measure (estimated using ICC between adjacent time windows for readout) increased with longer aggregation scopes. The metrics “number of daily bouts,” “bout duration, SD,” and “number of short bouts” reached an ICC ≥0.75 at the earliest aggregation scope (12 days). Ultimately, all digital measures achieved an ICC ≥0.75 by 16 days, which we defined as the potential time-to-reliability benchmark in the context of this study (Figure 4). ICCs reached a plateau at values ranging between 0.78 and 0.84, depending on the measure.
Figure 4. Time-to-reliability characterization: ICCs between adjacent readout windows according to aggregation scope duration, for the digital measures of interest. The line represents the ICC value plot, gray shading represents 95% CIs, and red annotations indicate the aggregation scope first exceeding an ICC value of 0.75. ICC: intraclass correlation coefficient. See Multimedia Appendix 3 for higher resolution image.

Discussion

This report expands upon prior research [30,31], presenting a comprehensive application of an algorithm that captures step count and other aspects of mobility, such as walking cadence and bouts. We characterized the accuracy and reliability of this comprehensive set of digital walking measures from users wearing a wrist-worn device in real-world environments. We showed that these measures of walking reached reliable readings at around 16 days of wear time, and their levels of agreement with the reference device, measured by ICC, ranged between 0.7 and 0.9, a performance that supports their deployment in clinical trial settings with confidence.

Mobility and walking behaviors represent meaningful aspects of health, known to be associated with quality of life in general and clinical prognosis in specific settings [33-35]. Therefore, improved methods to measure mobility and walking behaviors have the potential to improve clinical care and clinical trial efficiency. One of the goals of our research is to build accurate tools to objectively quantify the aspects of walking behavior and extract clinically meaningful information in discrete populations of interest. In prior work, we have developed algorithms whose outputs (measures for step counts and ambulatory time) demonstrated sensitivity to treatment effects in patients with Lewy body dementia [31]. We have also characterized the accuracy of an iteration of that algorithm in a cohort of diverse individuals in the real world [30].

To our knowledge, this report is the first to characterize the amount of data required to ensure that a digital measure is reliable in a real-world setting. We developed a novel analytic
approach to characterize time to reliability, that is, the time needed for a measure to reach a degree of stability. Time to reliability is an important consideration to inform the design of clinical studies tracking real-world data, as it relates to specific metrics of interest. In this study, the time to reliability overall for all measures was ≤ 16 days (ICC ≥ 0.75 between nonadjacent readouts for all measures at day 16, Figure 4). This study included healthy individuals; in a clinical context, we anticipate that the stability of any given measure over time will be dependent on the type and severity of the disease of interest. It is reasonable to speculate that a mostly healthy cohort may demonstrate more variability and a larger distribution of walking behaviors than a cohort with disease burden, and this necessitates further research.

Most importantly, the algorithms developed to quantify daily step count and measures related to walking cadence and bouts were found to be accurate (agreement between the readouts from the study device and a highly accurate reference device ranged between ICCs of 0.7 and 0.9 for all measures, Table 2). Our study approach captured the measures of walking behavior in a real-world setting, over multiple days, to closely resemble actual use cases.

Considering the exponential growth of research on wearable sensors and related devices in recent years, it is important to place the capabilities described in this report in that context. In this work, we incorporated several key innovative approaches to address shortcomings present in comparable studies evaluating interdevice agreement. Prior studies have used colocated investigational and reference sensors (eg, 2 wrist-worn devices). But because of the known potential errors associated with body placement when capturing walking-related data [36-39], colocation could be vulnerable to bias toward overestimating performance. Our approach sought to mitigate that by using a highly accurate but pragmatic and ankle-worn source for ground truth labels. Further, most studies have narrowly focused on step counts [10-16] for short time periods in controlled laboratory environments (eg, only a single day in real-world settings), or when investigating walking bout and cadence or pace measures, they had a limited scope, with small samples of less than 40 participants [40-42] and short tests (sessions lasting 1 hour or less) performed in clinic. Our study addresses these existing evidence gaps, presenting a set of digital walking measures that are comprehensive beyond step counts and characterizing their analytic performance (accuracy and reliability) extensively, with data accrued throughout multiday periods and in the course of daily living activities. Moreover, given the research heterogeneity (comparisons of different devices, different ground truth sources, and with different analytic approaches), any direct comparison of study results side by side has to be done with caution, which highlights the need for standardization noted in professional statements in this field [6,9,43,44].

This study had limitations in regard to the participant population and the performance quality thresholds. First, our cohort was limited in size and consisted of generally healthy participants. Future studies may be needed to characterize the generalizability of the performance of these measures in populations with particular kinetic hallmarks (eg, neurological conditions, stroke, and trauma) or with mobility capacity issues (eg, cardiovascular or respiratory conditions). Our approach to determine time to reliability can be applied across studies in any therapeutic area and can guide study design requirements for wear time compliance. One aspect that will require attention is the optimization of actual compliance with hypothetical protocol specifications about wear time because this is a device intended for daily use life (for instance, our reliability analysis filtered participant-day data based on a threshold of 12 hours of daily wear time for 50% of the days over an evaluation period, but we did so retrospectively). Second, although we report on performance parameters, the definition of an acceptable performance (accuracy and reliability) quality threshold remains undefined in the field. We did not prespecify performance categories in this study, but, for instance, prior accuracy studies have categorized agreement ICC values 0.7-0.9 as moderate to good [40,45], and reliability studies for patient-reported outcomes have considered test-retest ICC >0.5 as acceptable [46]. Importantly, what constitutes a clinically meaningful change for each of the measures of walking will likely depend on the therapeutic area under consideration. Further research is also needed to address which of these measures can provide clinically relevant insights in a given population. This set of digital walking measures has the potential to convey comprehensive information beyond flat quantification (via step counts), about aspects such as maximal walking capacity, endurance, or activity patterns during daily living, which may have different relevance or sensitivity to detect status changes depending on the health setting (for instance, cardiopulmonary conditions, oncology, or neurology). Furthermore, the optimization of the clinical utility of these measures may require their aggregation into composite metrics. The potential complexity of this future research brings to the forefront the importance of establishing first a thorough understanding of their individual analytical performance, which this study does. We believe that the accuracy and reliability results detailed here are the first step to support the use of digital measures of walking as feasible and reliable end points in clinical studies.

In conclusion, we have developed algorithms that accurately quantify daily step counts and measures of walking cadence and bouts from users wearing a wrist-worn device in a real-world setting. Further, we have also developed a novel method for characterizing the time required for a digital measure to stabilize (time to reliability). Given the growing use of wearable sensors to measure aspects of health, these findings may guide practical implementation of these digital measures of walking behavior into clinical studies.

Acknowledgments
The authors wish to acknowledge writing and editing support from Julia Saiz from Verily Life Sciences. This study was sponsored by Verily Life Sciences who was responsible for data collection. The authors had access to the raw study data in full.
Data Availability
Data from this study are not available due to the nature of this program. Participants did not consent for their data to be shared publicly.

Authors’ Contributions
NK, ER, and RK conceptualized and designed the study. Verily Life Sciences LLC collected the data. NK, SS, PB, and SP analyzed and interpreted the data. All authors interpreted the results, wrote and reviewed the manuscript, and approved the final manuscript for submission.

Conflicts of Interest
All authors report employment and equity ownership in Verily Life Sciences.

Multimedia Appendix 1
Additional methods and results.
[DOCX File, 101 KB - humanfactors_v10i1e48270_app1.docx ]

Multimedia Appendix 2
High resolution version of Figure 2.
[PNG File, 1229 KB - humanfactors_v10i1e48270_app2.png ]

Multimedia Appendix 3
High resolution version of Figure 4.
[PNG File, 881 KB - humanfactors_v10i1e48270_app3.png ]

References


Abbreviations

ICC: intraclass correlation coefficient

---

©Nathan Kowahl, Soooyoon Shin, Poulami Barman, Erin Rainaldi, Sara Popham, Ritu Kapur. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 03.08.2023. This is an open-access article distributed under the terms of the
Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Assessing Mood With the Identifying Depression Early in Adolescence Chatbot (IDEABot): Development and Implementation Study

Anna Viduani1,2, BA, MSc; Victor Cosenza1, BA; Helen L Fisher4,5, PhD; Claudia Buchweitz1,2, BA, MSc; Jader Piccin1,2, MD, MSc; Rivka Pereira1,2, BA; Brandon A Kohrt6, MD, PhD; Valeria Mondelli7, MD, PhD; Alastair van Heerden8, PhD; Ricardo Matsumura Araújo3*, PhD; Christian Kieling1,2*, MD, PhD

1Department of Psychiatry, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil
2Child and Adolescent Psychiatry Division, Hospital de Clínicas de Porto Alegre, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil
3Center for Technological Advancement, Universidade Federal de Pelotas, Pelotas, Brazil
4Social, Genetic & Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King’s College London, London, United Kingdom
5Economic and Social Research Council Centre for Society and Mental Health, King’s College London, London, United Kingdom
6Division of Global Mental Health, Department of Psychiatry, School of Medicine and Health Sciences, The George Washington University, Washington, DC, United States
7Department of Psychological Medicine, Institute of Psychiatry, Psychology, King’s College London, London, United Kingdom
8Centre for Community Based Research, Human Sciences Research Council, Pietermaritzburg, South Africa

*these authors contributed equally

Corresponding Author:
Christian Kieling, MD, PhD
Department of Psychiatry
Universidade Federal do Rio Grande do Sul
Rua Ramiro Barcelos, 2400
Porto Alegre, 90035003
Brazil
Phone: 55 5133085624
Email: ckieling@ufrgs.com

Abstract

Background: Mental health status assessment is mostly limited to clinical or research settings, but recent technological advances provide new opportunities for measurement using more ecological approaches. Leveraging apps already in use by individuals on their smartphones, such as chatbots, could be a useful approach to capture subjective reports of mood in the moment.

Objective: This study aimed to describe the development and implementation of the Identifying Depression Early in Adolescence Chatbot (IDEABot), a WhatsApp-based tool designed for collecting intensive longitudinal data on adolescents’ mood.

Methods: The IDEABot was developed to collect data from Brazilian adolescents via WhatsApp as part of the Identifying Depression Early in Adolescence Risk Stratified Cohort (IDEA-RiSCo) study. It supports the administration and collection of self-reported structured items or questionnaires and audio responses. The development explored WhatsApp’s default features, such as emojis and recorded audio messages, and focused on scripting relevant and acceptable conversations. The IDEABot supports 5 types of interactions: textual and audio questions, administration of a version of the Short Mood and Feelings Questionnaire, unprompted interactions, and a snooze function. Six adolescents (n=4, 67% male participants and n=2, 33% female participants) aged 16 to 18 years tested the initial version of the IDEABot and were engaged to codevelop the final version of the app. The IDEABot was subsequently used for data collection in the second- and third-year follow-ups of the IDEA-RiSCo study.

Results: The adolescents assessed the initial version of the IDEABot as enjoyable and made suggestions for improvements that were subsequently implemented. The IDEABot’s final version follows a structured script with the choice of answer based on exact text matches throughout 15 days. The implementation of the IDEABot in 2 waves of the IDEA-RiSCo sample (140 and 132 eligible adolescents in the second- and third-year follow-ups, respectively) evidenced adequate engagement indicators, with
Background

The challenges and limitations of the current tools of mental health assessment—mostly performed using standardized scales—have increased the interest in alternative monitoring tools. Traditional assessment often fails to incorporate the dynamic nature of psychological constructs and other relevant clinical features [1] and is not capable of capturing prognostic and therapeutic differences among patients [2] as well as the personalized aspects that are essential to address mental health issues.

Over recent decades, technology has created an opportunity to expand data collection and analysis beyond clinical and research facilities and centers, with flexibility to create participative, 2-way communication applications that can be easily adapted and used in everyday settings for a variety of target populations [3]. Considering the central role of language in the diagnosis and assessment of mental health, a shift toward a technology focused on conversational aspects may be key to systematizing natural language domains that are not currently explored in clinical settings [4].

In this sense, we propose that using chatbots—digital systems that rely on a conversational interaction that mimics human conversation [5]—may be an alternative to using traditional assessment methods. Chatbots are capable of capturing real-time accounts of events (ie, at the moment the event is being experienced) [6] and thus may further our current understanding of time- and context-contingent associations among activities, moods, and experiences [7]. Primarily, it has been theorized that chatbots both facilitate disclosure [8,9] and provide an opportunity to collect real-time information on mood and behavior in real-world settings with lower perceived burden for participants and researchers, increasing ecological validity, minimizing recall biases [10], and taking advantage of human-like conversation features to assess psychological constructs (such as depression) in a scalable, systematic fashion that is not possible with the usual application of instruments and scales.

One important advantage of chatbots is that they may be integrated into existing applications that are routinely used by the general public and designed as affordable, potentially scalable tools, following a frugal innovation model [11]. In addition, chatbots could be explored to reduce barriers that typically prevent identification of mental health disorders among, and help-seeking by, young people, a group especially susceptible to these conditions [12]. Given the scarcity of resources allocated to mental health care, particularly in middle-income countries such as Brazil, the development of frugal chatbot apps is a promising alternative.

Objectives

Chatbots have been used in mental health research for purposes such as therapy, training, and screening [13,14]. Nevertheless, most studies on user-chatbot interactions have focused on adults [15], although adolescents are often more familiar with smartphones than other populations [16]. Thus, exploring the feasibility of using chatbots to collect data on adolescent mood and behavior in an ecological fashion may be a promising avenue of inquiry. We hypothesize that, by leveraging already existing technologies, chatbots are a feasible, viable form of monitoring changes in mood and symptoms over time in adolescent populations. Moreover, we believe that their use lessens participant burden, possibly augmenting sustained engagement with the tool.

Therefore, we aimed to develop a chatbot tool to collect real-life data on mood and behavior from adolescents using text and audio messages. Here, we present the development and feasibility pilot of and initial results obtained with the implementation of the WhatsApp-based Identifying Depression Early in Adolescence Chatbot (IDEABot).

Methods

Study Setting: Identifying Depression Early in Adolescence Risk Stratified Cohort

The IDEABot was developed as part of the Identifying Depression Early in Adolescence Risk Stratified Cohort (IDEA-RiSCo) study [17]. The IDEA-RiSCo study includes 150 Brazilian adolescents (n=75, 50% female participants and n=75, 50% male participants) aged 14 to 16 years at baseline, stratified into 3 groups: low risk for developing depression (50/150, 33.3%), high risk for developing depression (50/150, 33.3%), and experiencing a current untreated major depressive episode (50/150, 33.3%). Participants were selected for each...
group using the Identifying Depression Early in Adolescence Risk Score (IDEA-RS), an empirically generated algorithm developed to estimate the individual-level probability of a unipolar depressive episode 3 years after initial assessment [17-19]. Additional details on procedures used in the IDEA-RiSCo study are described elsewhere [17].

**Rationale and Feasibility Pilot**

The IDEABot was developed to collect data from Brazilian adolescents via WhatsApp (Meta) [11]. In 2019, WhatsApp was reported to have been used at least once every hour by 81% of Brazilians [20]. Moreover, among adolescents from public state schools in the city of Porto Alegre, Rio Grande do Sul, Brazil (the population from which the IDEA-RiSCo sample was derived), WhatsApp was the most popular web-based platform, used at least once a day by 90% of the sample [21].

The IDEABot was devised to collect daily data on current mood via both structured items or questionnaires and free audio reporting of the aspects of daily life considered by participants (Multimedia Appendix 1). An interdisciplinary team was engaged in the project, including mental health practitioners (psychiatrists and psychologists), computer scientists, and writers. The prototype version of the IDEABot was designed and implemented in Brazilian Portuguese using inputs from the research team, followed by a feasibility pilot that generated a round of adjustments.

For the feasibility pilot, 6 adolescents were invited to test a prototype version of the IDEABot and comment on their user experience. They tested the chatbot system for 5 days, during which they answered the Short Mood and Feelings Questionnaire (sMFQ) and participated in 2 days of brief audio recordings. All features and possible response modes were tested. After test completion, the adolescents participated in an individual interview and a focus group discussion, conducted on the web by 2 researchers (AV and CK).

The interviews focused on the overall experience, feasibility, and acceptability of using the IDEABot (including concerns about data safety and privacy). In addition, the adolescents were engaged in jointly exploring and proposing improvements and solutions for perceived problems. In the focus group, anchored vignettes were used [22] to explore participants’ perceptions of the chatbot (Multimedia Appendix 2).

**Implementation of the Final Version of the IDEABot**

After the pilot test, the final version of the IDEABot was generated and subsequently implemented in the second- and third-year follow-ups of the IDEA-RiSCo study [17]. On the basis of a review of the literature, the following usability indicators [23] were evaluated to define successful implementation [24,25]: (1) acceptance (ie, the proportion of participants who were invited to take part in the IDEABot data collection and agreed to use the tool); (2) initial attrition (ie, failure to further engage in the protocol after agreeing to participate in the data collection and complete the initial steps); and (3) compliance, defined as the proportion of days on which participants generated at least 1 data point over the 15 days of data collection.

Socioeconomic status was also assessed with data collected at baseline using the Brazilian Criterion of Economic Classification [26], along with administration of a 9-item questionnaire on the frequency of the participants’ use of 8 social media platforms, including the frequency of WhatsApp use [21,27]. Responses were aggregated into 3 strata (1=never, 2=several times/week, and 3=several times/day or constantly).

Categorical and numerical variables were compared using the chi-square and Mann-Whitney U tests, respectively. In addition, the Spearman correlation coefficient was used to verify correlations among continuous variables. All analyses were performed using SPSS software (version 26.0; IBM Corp).

**Ethics Approval**

The development and research use of the IDEABot was approved by the Hospital de Clínicas de Porto Alegre ethics committee (50473015.9.0000.5327).

**Informed Consent and Participation**

All adolescents and caregivers provided written assent or consent to participate in each stage of data collection and were given the opportunity to withdraw assent or consent at any time. For participants aged >18 years, written consent was obtained directly. If participants wished to stop receiving messages from the chatbot before the completion of the 15-day trial, they were instructed to contact a research team member. In addition, participants were instructed to use the WhatsApp delete button if they preferred to delete sent messages or audio files. Along with the research team’s explanation on the functioning of the IDEABot, the chatbot’s first interaction with the user explicitly stated the nature of the exchange that would take place. Participants were thus aware that the audio recordings were not listened immediately and that the chatbot was not a channel for seeking help. Participants were provided with an additional telephone number and instructed to contact a team member (a board-certified psychiatrist) in case they were actively seeking information related to mental health issues. Furthermore, participants received information regarding the national helpline for health and safety emergencies. Following Brazilian legislation, participants did not receive financial incentives for taking part in the study but were offered compensation for mobile internet data use during their participation.

**Results**

**Results of the Feasibility Pilot**

Six adolescents (n=4, 67% male participants and n=2, 33% female participants; n=1, 25% of the 4 male participants had lived experience of depression, as did n=1, 50% of the 2 female participants) aged 16 to 18 years participated in the feasibility pilot. They were selected by convenience among the group of adolescents who had already participated in other projects conducted by our research team. Despite their heterogeneous socioeconomic backgrounds, all had a smartphone with internet access. Parental consent was obtained for all underage participants (those aged <18 years). As most of the participants (5/6, 83%) had already participated in other stages of the research, they were familiar with the investigators and knew about the IDEA-RiSCo objectives and procedures. The
Development of the IDEABot

The IDEABot was successfully developed to perform prescribed interactions requesting audio and text responses from participants to the questions it posed. The chatbot questions and responses were expressed only in text format, regardless of the format of user input. The IDEABot was also designed to delay answers proportionally to the length of the text being sent to users to simulate a more natural typed conversation. Using a rule-based approach, four types of interactions were developed: (1) mood ratings, (2) emoji mood ratings, (3) brief audio recordings, and (4) questionnaire answers (Multimedia Appendix 3).

As a first step to activate the chatbot, users were required to send a WhatsApp text message (any content was acceptable) to the chatbot’s mobile number. To ensure both the standardization of instructions given to users and clarity regarding the nature of the conversation, as well as to prevent misconceptions (such as participants believing that the chatbot is a real person or that the audio recordings would be listened immediately), the first interaction with the chatbot was designed to review overall functionality. This initial interaction was named day 0 and covered the routines that users should expect over the subsequent 14 days and how they were supposed to respond. Because of the IDEABot’s nature and objective, data generated on day 0 will be excluded from future analyses.

The chatbot follows a time-contingent sampling for each participant. In this sense, it is designed to initiate interactions at fixed times: every day, beginning at 1:30 PM, participants receive a message asking whether they are available to answer the scheduled questions. They may answer immediately after the first message prompt or use a snooze function to schedule a reminder for a later time in the day (the IDEABot allows snoozing until 3 AM the next day). If participants ignore the first prompt, additional messages are sent at 3-hour intervals. Participants have until 6 AM the following day to respond to the questions of each daily cycle. If the interaction is not completed, at 10 AM the following day, the chatbot informs the participant that the daily cycle will end without completion and that a new daily cycle will begin, also providing the time when the next message would be sent. In addition to scheduled interactions, participants are also given the option to send unprompted audio recordings throughout the day (Figure 1).

The chatbot’s schedule is divided into five interaction modes: (1) introduction (the first interaction with users), (2) audio questions, (3) administration of a version of the sMFQ, (4) other messages, and (5) the snooze function (Multimedia Appendix 3). On 7 (47%) of the 15 days, IDEABot asks broad questions about daily life, social interactions, and preferences (Textbox 1), and participants are invited to answer through audio recordings. The goal is to collect at least 1 minute of audio recordings per day from each participant. If the answers provided by participants to the 2 daily questions do not add up to 1 minute in duration, the chatbot asks 2 standard follow-up questions, encouraging the participant to say more. If after the first follow-up question (“Thank you for sending this audio! Tell us a little bit more about it, [participant]!”), the total interviews lasted 20 to 30 minutes, whereas the duration of the focus group was 50 minutes.

Overall, participants considered the IDEABot easy to use and enjoyable. All 6 adolescents completed at least 4 (80%) of the 5 interactions and sent an average of 54.5 (range 2-97) seconds of audio recordings per day. The adolescents expressed that directed questions (such as those asking about their daily routine) were easier to answer than more open questions (such as the initial request for participants to introduce themselves). In addition, the adolescents considered the prompts that targeted the collection of at least 1 minute of audio recordings over the day to be adequate.

Overall, they perceived the burden of integrating the chatbot into their daily routine as low. In fact, they highlighted a positive effect of talking about their daily lives:

It was a good experience...I felt I was talking about my things to someone—it even sounded like there was someone there wanting to know how my day was. Sometimes you spend your day without anyone asking you that. But the chatbot asked. [Female participant, aged 17 years]

Regarding the sMFQ, the adolescents found that some of the instructions provided by the chatbot were unclear and made suggestions on how to fix these issues. It asked participants to answer the sMFQ using the numbers 0, 1, or 2. The adolescents suggested further anchoring of these responses (eg, through reminders of the meaning of each number during the completion of the questionnaire). The instructions were adjusted accordingly after these difficulties and possible solutions were explored with the adolescents. In the final version, an explanation of each possible choice of answer was provided (0=no, 1=sometimes, and 2=yes) before the participants were asked to complete each item of the sMFQ, using, for example, the statement “I feel sad today.” In addition, a short reminder of the meaning of each numeric answer (0, 1, or 2) was added after each chatbot prompt.

An important adjustment made possible by the feasibility pilot was as follows: the adolescents tended to respond to the chatbot’s final interaction by either thanking it or sending an emoji. In the chatbot’s initial programming, this was interpreted as an unsolicited interaction to which the IDEABot responded by requesting an audio message to explain what the participant had said. This chatbot response would often confuse the adolescents. To avoid this, we developed a content-based rule: if participants responded with a predefined set of words (“ok,” “see you,” “thank you,” or variations), this was interpreted as a conversation closure, and the chatbot’s probe would not be triggered.

Another aspect that required changing was suggested by the adolescents in relation to the schedule of interactions. The adolescents argued that they would most likely be at school or asleep at 10:30 AM and therefore would probably not feel comfortable responding to the questions owing to their current environment (especially if they were at school). The adolescents then suggested that the first interaction of the day be moved from 10:30 AM to 1:30 PM, which was implemented in the final version of the IDEABot.
duration of the audio recording still does not reach 1 minute, the chatbot sends the second question (“It would be very important if you could tell us a little more, okay?”). Regarding this last question, participants can choose whether to send another audio recording (typing “yes” or “no” before sending the audio recording). One example is provided in Figure 2.

On the 7 days without audio prompts, participants are asked to complete the sMFQ [28,29]. The 13 questions of the sMFQ cover the current day (instead of the last 2 weeks as in the original sMFQ; Multimedia Appendix 4). Participants are instructed to type 0, 1, or 2 to answer each question, and they have the option to correct their answers (for relevant aspects of the processing of the collected data and analyses, refer to Multimedia Appendix 5 [30-36]).

Figure 1. Overview of the functioning of the Identifying Depression Early in Adolescence Chatbot over the period of a day.
### Textbox 1. Questions (question 1 [Q1] and question 2 [Q2]) or prompts for audio responses requested by the Identifying Depression Early in Adolescence Chatbot (the original questions are in Brazilian Portuguese).

<table>
<thead>
<tr>
<th>Day 1</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Q1. Can you introduce yourself?</td>
<td></td>
</tr>
<tr>
<td>• Q2. What have you done today? Is your day going according to your usual routine?</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Day 3</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Q1. Are you at home?</td>
<td></td>
</tr>
<tr>
<td>• [If the response is “yes”] What are you doing? Is someone else around?</td>
<td></td>
</tr>
<tr>
<td>• [If the response is “no”] Who do you live with? Do you get along with the people you live with?</td>
<td></td>
</tr>
<tr>
<td>• Q2. Can you tell me more about your house? Do you like living there?</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Day 5</th>
<th>Questions</th>
</tr>
</thead>
</table>
| • Q1. Did you go outside today at all, [participant]?
| • Do you spend more time inside, or do you sometimes go out? When you’re out, what do you normally do? |
| • Q2. And how’s your neighborhood? Are there nice things around? |

<table>
<thead>
<tr>
<th>Day 7</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Q1. Today I want to know about your favorite story. What is it? You can choose a movie, a series, a book...whatever you want!</td>
<td></td>
</tr>
<tr>
<td>• Q2. And why is this your favorite story, [participant]?</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Day 9</th>
<th>Questions</th>
</tr>
</thead>
</table>
| • Q1. Do you use your mobile phone a lot, [participant]?
| • What are your favorite things to do on the mobile phone? |
| • Q2. And how much time do you think you spend on the internet each day? Do you use the internet mostly during the day or at night? Why? |

<table>
<thead>
<tr>
<th>Day 11</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Q1. Not counting the audio recordings you send here [grinning face with sweat emoji], who do you talk to about things that happen in your life? How’s your relationship with this person?</td>
<td></td>
</tr>
<tr>
<td>• Q2. And why do you trust this person?</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Day 13</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Q1. It’s been almost 2 weeks since we started talking, [participant]! How did you feel about answering these questions?</td>
<td></td>
</tr>
<tr>
<td>• Q2. And how have you been in these last 2 weeks? Has anything different happened?</td>
<td></td>
</tr>
</tbody>
</table>
Initial Results of a Full-Sample Implementation of the IDEABot

The IDEABot was first implemented as part of the IDEA-RiSCo second-year follow-up assessment, which took place between August 1, 2020, and January 31, 2022. It was subsequently also used in the third-year follow-up of the IDEA-RiSCo sample, which occurred between August 1, 2021, and September 30, 2022.

To explain the chatbot's functioning and features to participants, an animated video (Multimedia Appendix 6) was developed by the research team, providing a comprehensive overview of the research process. It reminded participants about the previous waves of data collection and the overall research goal, as well as presented the various steps of data collection that they could engage in (including the IDEABot). In addition, the video provided information regarding data confidentiality, including end-to-end encryption by WhatsApp for all chats, and the...
measures taken by the research team to ensure data protection. After the video was sent, if participants agreed to use the IDEABot, a research team member sent a link that directed users to initiate the interaction.

For the second- and third-year follow-up assessments, 9.7% (11/113) and 11.5% (14/122) of the adolescents, respectively, did not have a smartphone and agreed to receive a device from the study team to enable data collection completion. All other participants used their own smartphones and already had WhatsApp installed. In terms of technical challenges experienced during the IDEABot implementation, we recorded 6 and 14 occurrences or technical malfunctions in the second- and third-year follow-up assessments, respectively.

In the second-year follow-up, there were 5 issues with the integration with WhatsApp’s application programming interface (API; September 11 and 15, 2020; December 11, 2020; April 4, 2021; and June 15, 2021) and 1 instance in which WhatsApp was offline around the world owing to an instability in Meta’s servers (October 4, 2021) [37]. All issues were resolved within 24 hours, but the interactions of 6.2% (7/113) of the participants were affected directly. As result, these participants lost 8 interaction days in total. In addition, in the second-year follow-up, there were 3 instances in which the chatbot’s malfunctioning prevented participants from completing the scheduled interactions. In all cases, participants repeated the interaction days affected. Finally, there was 1 occasion on which a participant was not able to complete the day’s interaction owing to a problem with telephone billing, which was later resolved.

In the third-year follow-up, there were 12 issues with the integration with WhatsApp’s API (March 18 and 19, 2022; April 5 and 20, 2022; May 5, 18, and 20, 2022; June 14 and 26, 2022; July 8, 2022; and August 16 and 28, 2022), as well as 2 instances in which chatbot was unable to access the internet (October 10, 2021, and February 18, 2022). In addition, the instance in which WhatsApp was offline worldwide (October 4, 2021) also affected the third-year follow-up. Only 1 occurrence was not resolved within 24 hours (March 18 and 19, 2022), owing to the API’s instability. Interactions were affected for 33.6% (41/122) of the participants, resulting in a loss of 16 occasions in which these participants could have completed the day’s interaction. The greatest number of occurrences were mostly caused by the changes in WhatsApp Web, the web-based interface for WhatsApp required for running the API.

In the second-year follow-up, 140 adolescents took part in some aspects of data collection and were therefore eligible to use the IDEABot. Of the 140 adolescents, 113 (80.7%) agreed to use the IDEABot and completed the initial interaction. Of these 113 participants, 1 (0.9%) interacted with the chatbot only on the first interaction. The 112 adolescents who continued interacting with the chatbot engaged on average 12.8 (SD 3.5) of the 14 possible days, corresponding to a compliance rate of 91.4%. The snooze function was used 609 times, resulting in 331 completed interactions. In addition, participants sent on average 65 (SD 37.7) seconds of audio recordings per day, resulting in an average of 7.6 (SD 4.3) minutes of audio recordings per participant.

For the third-year follow-up, 132 adolescents took part in some aspects of data collection and were therefore eligible to use the IDEABot. Of the 132 adolescents, 122 (92.4%) agreed to use the IDEABot and completed the initial interaction. Of these 122 participants, 1 (0.8%) interacted with the chatbot only on the first interaction. The 121 adolescents who continued interacting with the chatbot engaged on average 10.57 (SD 3.4) of the 14 possible days, corresponding to a compliance rate of 75.5%. The snooze function was used 569 times, resulting in 258 completed interactions. In addition, participants sent an average of 69.2 (SD 66.1) seconds of audio recordings per day, resulting in an average total of 8.1 (SD 7.8) minutes of audio recordings per participant.

No significant association between socioeconomic status and the number of days of interaction with the IDEABot was found (P.88); the number of days on which responses were recorded also did not differ when participants were stratified according to the pattern of previous WhatsApp use (ie, never, several times/week, or several times/day; P.98) or by sex (male or female; P.66).

Discussion

Principal Findings

This study outlines the development, feasibility pilot, and initial results obtained with the implementation of a chatbot to support mood assessment in adolescents. Although chatbots are becoming increasingly more common in health care settings [38], few studies have provided detailed analyses and empirical discussions of specific design elements and development techniques [39]. In this sense, we believe that reporting the development and implementation of the IDEABot is a novel and relevant contribution, especially given the overall good acceptance for using the tool, low attrition, and high compliance in terms of the proportion of responses in relation to the total number of elicited prompts.

To the best of our knowledge, the IDEABot is the first chatbot specifically tailored to aid multimodal research data collection with adolescent populations. Our decision to use an existing platform made it possible to design, develop, and implement the IDEABot in a way that directly addresses the constraints that the use of new mobile apps may pose to research teams and users, in addition to saving development and adjustment time. The IDEABot runs on any smartphone with WhatsApp, regardless of operating system, as long as internet connectivity is available. The IDEABot thus qualifies as a frugal innovation: it is significantly cheaper than other alternatives (such as the development of a new stand-alone app); it has proven sufficient for the proposed level of data collection; and by using it, we were able to reach participants who would otherwise remain underrepresented [11]. Moreover, the proposed approach to data collection is highly flexible and could potentially leverage all forms of interactions available on WhatsApp, including photographs and video recordings.

The initial administration of the IDEABot indicates engagement rates of >80%, with more than half of the participants (59/113, 52.2% and 52/122, 42.6%) for second- and third-year follow-up.
use, respectively) completing all 15 days of collection. In ecological momentary assessment studies (ie, studies that are designed to collect individual data at several time points), 80% has been proposed as an indicator of adequate compliance [40]. Although compliance tends to vary in ecological momentary assessment studies (also depending on the number of measures made over time) [41], we believe that the rate obtained with the IDEABot matches the expected rates in similar studies and is adequate, considering the target population and that no financial or other direct incentive was used.

In this sense, we believe that repurposing an already ubiquitous tool in the life of adolescents to collect research data can increase overall engagement as well as diminish the perceived burden of data collection. Moreover, we highlight the importance of youth participation in the creation, adaptation, and implementation of the IDEABot. A chatbot’s personality, interaction flow, conversation length, and dialogue structure are important aspects and can influence user satisfaction [39].

In the case of the IDEABot, all these aspects were created and tailored with the aid of a group of adolescents, who were active in pointing out any strangeness or discomfort and were ready to brainstorm solutions. Thus, not only was the final chatbot tailored to collect relevant research data, but it was also pleasant in terms of appearance and the manner of interaction with adolescents themselves, which can greatly decrease the burden of research participation.

All things considered, the IDEABot still has important limitations that need to be addressed. Despite good engagement rates among Brazilian adolescents, the IDEABot is a basic chatbot that uses a rule-based approach. Although this gives the researchers optimal control over conversation flow and topics, the limited response range may decrease usability by adolescents (who may, for example, become frustrated with repeated error messages) [42]. In addition, as a WhatsApp-based chatbot, the IDEABot is susceptible to changes in policies and bugs affecting the platform. In this sense, the usability of the IDEABot becomes heavily linked to WhatsApp as a commercial product, and researchers have no control over policies such as data security and other features. The instance in which WhatsApp was offline worldwide preventing data collection is also an indication of the bot’s susceptibility to the platform’s functioning, which may hinder its applicability.

Furthermore, although the chatbot’s user-oriented design may contribute to higher self-disclosure [43], privacy concerns regarding the use of the data are a relevant topic. WhatsApp policies include “end-to-end encryption” [44], and the IDEABot also stores information (audio recordings and conversation logs) on secure encrypted servers with additional anonymization of sensitive information in reports. However, all conversation logs and sent audio files remain accessible to other users in the mobile phone or any other devices that may be used to connect to WhatsApp (such as WhatsApp Web). Local backups may also store this information in user’s mobile phones, creating the risk of confidentiality breaches that cannot be controlled by the research team.

Another important aspect is the chatbot’s response to serious health concerns. As the IDEABot often queries participants on mood and daily events, we might expect sensitive information to be disclosed at the moment when distressing events occur. However, the IDEABot’s rule-based approach may not be suitable for fully and effectively responding to these events. In our project, mitigation efforts included full disclosure that audio messages would not be listened to immediately by the research team and that the IDEABot was not equipped to deal with mental health emergencies. Participants were also provided with the national emergency service hotline number for acute cases, and they were also able to call a research team psychiatrist in case of significant distress during the data collection process. However, this particular safety measure was never used by participants during the data collection process in either follow-up wave.

Also important is the susceptibility of the interface to technical error, such as bugs in the chatbot response routine (it does not respond, or it provides responses that do not fit the conversation context). As people may anthropomorphize chatbots [43], perceiving them as having a mind with intention, consciousness, and goals [45], these instances may generate negative feelings or distress responses, with a potential negative impact on participants who could become attached to the chatbot [46], or even hinder retention and continuous use. For the IDEABot, preventive measures include continuous function supervision by both humans and software monitoring the integration with WhatsApp’s API. In addition, using the platform as a medium for data collection also gives researchers little control over the quality of the data while they are being collected. This can be critical, for example, during data analysis, in which the selection, extraction, and assessment of acoustic features are dependent on the quality of the audio files and the data obtained [30]. This highlights the need for further research to explore the data collected as well as the techniques that are best suited for collecting and analyzing the data.

Therefore, the IDEABot presents limitations that may be considered inherent to the methods chosen. However, its development was guided by the principle of user transparency, and challenges regarding privacy and adverse incidents have been, and continue to be, closely and continuously assessed throughout development, implementation, and use. In addition, we believe that, as a tool, the IDEABot supports stakeholder values [47]. Nonetheless, the ethical considerations involving chatbot use will change with time and technical development, and continuous reassessment is vital to address any resulting ethical concerns that may arise.

Conclusions

The IDEABot is a novel WhatsApp chatbot developed to aid intensive longitudinal collection of mood data among adolescents. The collection of audio recordings and information on mood and behavior throughout 15 days may enable analyses of adolescents’ data that would otherwise not be possible. The completion rate shows that the IDEABot was able to collect information in a manner that is attuned to the adolescents’ lives. In this sense, the use of sequenced audio recordings may be considered similar to an audio diary, capturing much of the sense making and representation of experiences at different time points [48].
It is worth noting that the choice for a multimodal data collection approach that combines audio recordings of prompted speech, daily information on mood, and traditional assessment methods (such as questionnaires) sheds light on aspects of depression—such as the temporal evolution of symptomatology—that have only recently become a focus of research and are also rapidly advancing. Thus, the IDEABot generates a rich database that combines different types of input information that can be compared and triangulated. The IDEABot is a frugal innovation and therefore has a goal to meet the basic needs of a population that would otherwise remain underserved [11]; a strength of the IDEABot is its reliance on an available ubiquitous medium as a way to reach a population that is still underrepresented in research [49,50]. However, adaptability is key, and thus we chose to use a simple rule-based approach, allowing the IDEABot to be easily implemented, both technically and economically. As a result, the IDEABot is a feasible tool for data collection that can be adapted, tested, and implemented in different settings and for different purposes.

Another strength of the IDEABot is its capability for intensive data collection over extended periods within a longitudinal 3-year research project with a careful phenotypic characterization of the sample, including multiple informants. Such intensive and momentary data collection can elucidate aspects of the overall trajectory of different groups of individuals, such as those taking part in the IDEA-RiSCo study. This group approach can be useful for monitoring change and fluctuations in mood and to address the overall trajectories of different groups over time. In addition, periods of intensive data collection in individual participants may capture unique changes or symptom fluctuation patterns that would not otherwise be detected [7], contributing important information regarding symptom connectivity and centrality over time. The contrast between group and idiographic findings provides a further level of information not usually available in traditional research designs. In this sense, in addition to furthering our understanding of individual and group trajectories, the characterization of the sample also provides an opportunity to further explore the patterns of chatbot-assisted data collection.

In summary, the initial apps of the IDEABot were successful. The IDEABot seems to be a feasible, potentially scalable tool to collect data that can further our understanding of how mood changes and develops over time among adolescents.

Acknowledgments
The authors are extremely grateful to the schools and individuals who participated in this study and to all members of the Identifying Depression Early in Adolescence (IDEA) team for their dedication, hard work, and insights. This project was supported by the Royal Academy of Engineering under the Frontiers Follow-On Funding scheme (FF19201161). The original IDEA project was funded by an MQ Brighter Futures grant (MQBF/1 IDEA). Additional support was provided by the UK Medical Research Council (MC_PC_MR/R019460/1) and the Academy of Medical Sciences (GCRFNG_100281) under the Global Challenges Research Fund. This work was also supported by research grants from Brazilian public funding agencies Conselho Nacional de Desenvolvimento Científico e Tecnológico (477129/2012-9 and 445828/2014-5), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (62/2014), and Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (17/2551-0001009-4). CK is a Conselho Nacional de Desenvolvimento Científico e Tecnológico researcher and an Academy of Medical Sciences Newton Advanced Fellow. CK and BAK are supported by the US National Institute of Mental Health (R21MH124072). HLF was partly supported by the Economic and Social Research Council Centre for Society and Mental Health at King’s College London (ES/S012567/1). VM was supported by the National Institute for Health and Care Research Maudsley Biomedical Research Centre hosted by South London and Maudsley NHS Foundation Trust and King’s College London and MQ funding (MQBF/4). The views expressed are those of the authors and not necessarily those of the funders, the National Health Service, the National Institute for Health and Care Research, the Department of Health and Social Care, the Economic and Social Research Council, or King’s College London.

Conflicts of Interest
VM has received research funding from Johnson & Johnson, but the research described in this paper is unrelated to this funding. All other authors declare no other conflicts of interest.

Multimedia Appendix 1
Technical aspects of the development of the Identifying Depression Early in Adolescence Chatbot (IDEABot).
[DOCX File, 14 KB - humanfactors_v10i1e44388_app1.docx ]

Multimedia Appendix 2
Anchoring vignettes.
[DOCX File, 2390 KB - humanfactors_v10i1e44388_app2.docx ]

Multimedia Appendix 3
The types of interactions users can have with the Identifying Depression Early in Adolescence Chatbot (IDEABot).
Multimedia Appendix 4
Chatbot script for the Short Mood and Feelings Questionnaire instructions.

Multimedia Appendix 5
Processing and analysis of data and questionnaires.

Multimedia Appendix 6
Animated video.

References


Abbreviations

API: application programming interface
IDEA-Bot: Identifying Depression Early in Adolescence Chatbot
IDEA-RiSCo: Identifying Depression Early in Adolescence Risk Stratified Cohort
IDEA-RS: Identifying Depression Early in Adolescence Risk Score
sMFQ: Short Mood and Feelings Questionnaire
Perceived Use Cases, Barriers, and Requirements for a Smart Health-Tracking Toilet Seat: Qualitative Focus Group Study

Sander Hermsen¹, PhD; Vera Verbiest¹, MSc; Marije Buijs², MSc; Eva Wentink¹, PhD

¹OnePlanet Research Center, Wageningen, Netherlands
²Radboud University, Nijmegen, Netherlands

Corresponding Author:
Sander Hermsen, PhD
OnePlanet Research Center
Bronland 10
Wageningen, 6708 WH
Netherlands
Phone: 31 317 791 009
Email: sander.hermsen@imec.nl

Abstract

Background: Smart bathroom technology offers unrivaled opportunities for the automated measurement of a range of biomarkers and other data. Unfortunately, efforts in this area are mostly driven by a technology push rather than market pull approach, which decreases the chances of successful adoption. As yet, little is known about the use cases, barriers, and desires that potential users of smart bathrooms perceive.

Objective: This study aimed to investigate how participants from the general population experience using a smart sensor-equipped toilet seat installed in their home. The study contributes to answering the following questions: What use cases do citizens see for this innovation? and What are the limitations and barriers to its everyday use that they see, including concerns regarding privacy, the lack of fit with everyday practices, and unmet expectations for user experience?

Methods: Overall, 31 participants from 30 households participated in a study consisting of 3 (partially overlapping) stages: sensitizing, in which participants filled out questionnaires to trigger their thoughts about smart bathroom use and personal health; provotyping, in which participants received a gentle provocation in the form of a smart toilet seat, which they used for 2 weeks; and discussion, in which participants took part in a web-based focus group session to discuss their experiences.

Results: Participants mostly found the everyday use of the toilet, including installation and dismantling when necessary, to be relatively easy and free of complications. Where complications occurred, participants mentioned issues related to the design of the prototype, technology, or mismatches with normal practices in using toilets and hygiene. A broad range of use cases were mentioned, ranging from signaling potentially detrimental health conditions or exacerbations of existing conditions to documenting physical data to measuring biomarkers to inform a diagnosis and behavioral change. Participants differed greatly in whether they let others use, or even know about, the seat. Ownership and control over their own data were essential for most participants.

Conclusions: This study showed that participants felt that a smart toilet seat could be acceptable and effective, as long as it fits everyday practices concerning toilet use and hygiene. The range of potential uses for a smart toilet seat is broad, as long as privacy and control over disclosure and data are warranted.

(JMIR Hum Factors 2023;10:e44850) doi:10.2196/44850

KEYWORDS
digital health; internet of things; human factors; health tracking; device; automated; biomarker; personal health; personal hygiene; hygiene; data; privacy; innovation; mobile phone

Introduction

Background
With the rapid development of sensor technology and machine learning, novel opportunities for the unobtrusive and continuous detection of health issues have arisen. These opportunities have the potential to improve the prevention and treatment of debilitating health conditions through early detection and exacerbation signaling while also reducing patient burden by making invasive testing redundant. In theory, almost every
object surrounding people in daily life could be transformed into a smart entity by equipping it with sensors, actuators, and algorithms for the automatic evaluation of generated data. One promising area where unobtrusive and continuous detection can lead to great health benefits is the toilet. First, the toilet is a location where nearly everybody spends time regularly. Second, the toilet offers unrivaled opportunities for the automated measurement of a range of biomarkers and other data. The consistency, color, and density of urine, for instance, could offer insights into water-loss dehydration [1], a condition that occurs in 20% to 30% of the older population [2]. Ketones in urine are useful for detecting type II diabetes [3,4], a condition that affects >500 million people worldwide, with prevalence expected to grow even further in the next 10 years [5]. Detecting albumin and creatinine in urine can shed light on kidney failure [6]. Urine and stool contain proteins and leukocytes, which can provide information on the prevalence of inflammatory bowel disease [7], which exceeds 0.3% of the population in North America, Oceania, and most of Europe [8]. Similarly, many people have, or are at risk of, debilitating conditions associated with high blood pressure, which can also be measured during a bathroom visit, for instance, through strain detection [9].

With such great potential for automatic, unobtrusive assessment of relevant biomarkers, it may be no surprise that there have been several recent initiatives to develop such a smart toilet (eg, the studies by Wang and Camilleri [10], Bhatia et al [11], Bae and Lee [12], and Balaceanu et al [13]). These initiatives are mostly driven by a technology push rather than market pull approach: scientific and technological innovations serve as the drivers of solutions to societal problems rather than direct demand from a customer or an envisioned target population [14,15]. Technology push approaches play a major role in innovation, both by providing solutions for gaps between the status quo and desired societal states and by enabling new modes of idea generation and selection [16]. However, this approach is not without risks. A limited connection to people’s goals and barriers often leads to nonimplementation [17]. Furthermore, when implemented, 80% of newly introduced inventions fail within 2 years [18].

A major factor determining the success of technology push–driven innovations is the consideration of the barriers and needs of potential end users [15,18], with unmet demands and needs known to significantly impact the sustained use of digital tracking devices [19]. Traditional ways to incorporate user needs and demands into the development process are user experience research, which evaluates users’ opinions regarding the aesthetic, hedonic, affective, or experiential aspects of the use of a given technological prototype [20], and user-centered design, a methodology for placing users at the center of the development process from the early stages of designing system requirements to implementing and evaluating the product [21]. Although there is a definite value in having potential users of an innovation take part in the development process, this involvement is the most valuable when at the very conception of the innovation [22].

Recently, there have been new developments in early user involvement in scientific research under the guise of extreme citizen science, a participatory research approach in which citizens not only take part in gathering data but also codetermine the research agenda. Typically, when the term “citizen science” is used to describe a scientific work, it indicates that nonprofessional researchers gathered and occasionally processed data as part of the larger research endeavor. The widespread use of information and communication technology in general and ubiquitous computing in particular; the understanding that the public can supply free labor, skills, computing power, and even funding (crowdsourcing and crowdfunding); and the rising expectations of research funders for public engagement are all significant driving forces behind the recent growth in citizen involvement in research [23,24]. A relatively new development, however, is the development of extreme participatory approaches, in which citizens’ needs not only inform the development of an innovation but also determine the research questions that set the research agenda for the intervention development in the first place.

Unfortunately, how citizens can play this role often remains elusive. Turning everyday health challenges into research questions requires knowledge and skills that many people lack. Therefore, this study uses a citizen science approach based on participatory design methodology [25,26] to support citizens in capturing the potential use cases, user needs, and perceived barriers for smart sensor technologies in the bathroom. This methodology helps participants think about their situation and the ways in which technological innovations can or cannot support them in managing health conditions and living their everyday lives.

Although the discussion phase is exploratory and open to any input participants may provide, literature can already elucidate some of the themes that are likely to arise when thinking about, or trying out, smart bathroom technology innovations. First, the literature can shape one’s expectations of how smart bathroom innovations interact with everyday practices; these practices can be thought of as the interplay of practical knowledge, common understandings, rules, and material infrastructure that determines our expectations and behaviors at certain moments and places [27]. How the smart bathroom fits with people’s knowledge, common ideas and norms, and expectations surrounding toilet use determines the way in which it will be accepted, rejected, or even subverted for other use by future users [28]. This not only sheds light on the feasibility of the innovation but can also inform the design of future iterations of the innovation prototype. Second, literature on the use cases of technological innovations in everyday life shows that people have different uses for tracking technology, including directive tracking aimed at behavioral change; documentary tracking aimed at finding out more about oneself; diagnostic tracking aimed at answering questions about one’s health; tracking aimed at collecting rewards; and so-called fetishized tracking, that is, using technology out of love for the technology itself [29,30]. Third, the literature shows that sensing technology introduced in sensitive domains of everyday life triggers different privacy needs in different people [31]; some people are willing to use the technology out in the open and even use it to strengthen their social identity, whereas others are more reserved or hide the technology from others altogether—often referred to in the literature as on-stage use, off-stage use, and backstage use of...
technology, respectively [32]. Finally, data sharing needs, perceived barriers to sharing data, and privacy requirements differ between people [33], with some people being more trusting and willing to share data than others.

**Goal of This Study**

This study, therefore, aimed to investigate how participants from the general population experience using a smart toilet seat equipped with sensors for body temperature, weight, electrocardiogram, bioimpedance, and photoplethysmography installed in their home. The study contributes to answering the following questions: What use cases do citizens see for this innovation? and What are the limitations and barriers to its use in everyday life that they see?

**Methods**

**Overview**

The aim of this study was to investigate how participants from the general population experience using a smart toilet seat installed in their home and what use cases they foresee for such a toilet. The study consisted of 3 stages: sensitizing, provotyping, and discussion. The sensitizing [34] stage aims to help participants think about different aspects of the innovation. To do so, participants generally read materials, watch film clips, keep diaries, or fill out questionnaires that help them notice aspects and form their thoughts. In this study, participants filled out questionnaires to trigger their thoughts about smart bathroom use and personal health. The provotyping stage, a combination of “provocation” and “prototyping” [35,36], lets participants work with prototypes, often with low fidelity, of the innovation as a safe, gentle provocation. This helps elicit tacit knowledge such as everyday practices, norms, cultural conventions, and taboos. In this study, the prototype participants used a smart toilet seat for 2 weeks. The third stage is discussion, in which scientists and citizens explore themes and solution spaces together, based on the insights gathered in the sensitizing and provotyping stages. The participants took part in a web-based focus group session to discuss their experiences. In this study, the sensitizing and provotyping stages mostly overlapped. The recordings of the focus group sessions were transcribed and analyzed using thematic qualitative analysis.

This study was part of a larger trial testing the efficacy of sensors installed in the toilet: electrocardiogram sensors, bioimpedance sensors, photoplethysmography sensors, weight sensors, and body temperature sensors. The trial tested whether the sensors delivered adequate data quality to inform measurements and predictions and whether the data from the sensors enabled distinction between the different users of the toilet. We could not guarantee that the quality of the data was sufficient to provide valid and reliable feedback on biomarkers to participants. Furthermore, the data provided by the sensors did not contribute to the answering of the research questions in this paper. Therefore, participants received no feedback from sensor data of any kind, nor was the analysis and reporting of the sensor data part of this paper.

**Participants**

**Overview**

We aimed to include people from the general population, aged ≥16 years, and potentially interested in using a smart toilet. Participants were recruited from the province of Gelderland in the Netherlands and its neighboring regions owing to logistic restrictions in delivering and installing the toilet. To capture a potentially broad range of potential use cases, we aimed to include participants from all age groups, except children aged <16 years who could use the toilet as part of a participating household but could not actively participate and provide data. People weighing >100 kg were excluded from the study, as well as people with pacemakers and pregnant women, because the smart toilet prototype had not yet been tested for use with these groups. Because of the exploratory nature of the research, which aimed at generating use cases from a large populace and not specific groups, we added no further inclusion or exclusion criteria.

Because most potential participants lived in a household consisting of >1 person and the toilet seat collects data from every person using it, all members of the partaking households needed to give their consent to the collection of their physiological signals via the toilet seat. Therefore, we set up 2 levels of participation: active participation, in which the participant filled out all questionnaires and took part in the discussion session, and passive participation, in which the participant used the smart toilet but did not want their data to be used in the analysis. Passive participants did not fill out any questionnaires and did not take part in the discussion, and their physiological data were deleted after the measurement period. Only data from active participants were included in this study.

**Recruitment**

Participant recruitment took place through various publications in regional media, such as local newspapers and web-based news sites, and social media. Participants could indicate their interest by sending an email to the study coordinator, who then contacted them via email to inform them about the study procedure, aims, and time frame and share the consent form. If participants had any questions, the study coordinator answered them via email or telephone. If participants then agreed to take part, they filled out the consent form upon the delivery and installation of the smart toilet seat. Participants received no monetary or other remuneration for taking part.

**Sample Size Considerations**

In qualitative research, a priori sample size calculations are subject to conceptual debate and practical uncertainty. Saturation, that is, the moment when adding more data does not lead to new insights, is often seen as a criterion for the inclusion of more participants once the analysis has started. As a rule of thumb, 20 to 40 participants are usually considered sufficient to achieve saturation [37,38]. Given these considerations and the possibility of withdrawal, we aimed to recruit participants from 30 to 40 households for this study, with at least 1 participant per household. To ascertain a broad range of potential use cases, we aimed to recruit people from different age brackets, preferably >5 participants aged 16 to 30 years, >5
participants aged 31 to 45 years, >5 participants aged 46 to 65 years, and >5 participants aged >65 years.

Ethical Considerations
This study was deemed exempt from ethics approval according to the Dutch Medical Research Involving Human Subjects Act (Wet Medisch Onderzoek) by the medical ethical committee of the Maxima Medical Center in Veldhoven, Netherlands (decision number N21.090). An extensive risk assessment was performed and did not reveal any risks exceeding the acceptable limits, and possible risks were mitigated as much as possible. This study fully adhered to the Declaration of Helsinki, 2013 amendment.

Consent to Participate
All active and passive participants provided full written consent for their participation and the use of their data for scientific publishing and other dissemination purposes. Participants were briefed about the procedure and goal of the study and were aware that they could leave the trial at any point in time if they wished to do so without any consequences or obligation to give a reason.

Procedure and Materials
Overview
Upon the confirmation of participation, the research team sent out an information leaflet with general information; the goal, procedure, and background of the research; eligibility criteria; privacy considerations; and procedures for withdrawal and consent forms. They then made an appointment to deliver the smart toilet seat to the home of the participants. During the visit, all participants, both active and passive, signed the informed consent forms. Consent for participants aged <18 years was provided by their parents.

Sensitizing Phase: Questionnaires
Shortly after the installation of the toilet seat, all active participants filled out a web-based questionnaire on their mental well-being, gut health, overall health, and expectations toward the smart bathroom. To do so, they received an email containing an invitation link to the questionnaire, which was delivered through a web-based questionnaire delivery service (Castor EDC) and filled out on the participants’ own laptop, tablet, or smartphone. After this, participants received an email link to a second questionnaire, also delivered through Castor EDC, with questions regarding the installation clamps. During the 2-week use period, every evening at 7 PM, all active participants received an invitation to fill out a brief questionnaire via an ecological momentary assessment (EMA) app, which they had to install on their smartphone to participate in the study. To reduce their burden, participants were free to fill out or ignore the EMAs after filling out at least 4 of them during the 2-week study period to support the linking of sensor data to particular active users (not covered in this paper). The EMA questionnaires polled participants on toilet use but also contained 1 question each about general health, mood, and stress level and room to leave thoughts and questions about the smart toilet.

Finally, after the 2-week use period, active participants received an email invitation to a final questionnaire, which polled them about their experience using the smart toilet. This questionnaire contained questions from the Systemic Usability Scale (SUS) [39]. The SUS consists of 10 questions with a 5-item Likert scale ranging from “strongly disagree” to “strongly agree.” Because this scale has known limitations [40], additional items regarding the hedonic and pragmatic qualities of the prototype [41,42] were added. The hedonic quality of the prototype, which was measured on a 7-point Likert scale ranging from “strongly disagree” to “strongly agree,” corresponds to its valence and perceived usefulness, for example, its practicality, niceness, modernity, amusingness, credibility, ease of use, level of answering to needs, beauty, and robustness, and to disadvantages associated with its use, for example, intrusiveness, embarrassment, and nuisance. Pragmatic quality, measured on a similar 7-point scale, corresponds to the prototype’s perceived validity and reliability, for example, its exactness, level of detail, clarity, and credibility. Participants then filled out the Affinity for Technology Interaction (ATI) scale [43]. This questionnaire assesses a person’s tendency to actively engage in or avoid intensive technology interaction and consists of items measured on a 6-point Likert scale ranging from “completely disagree” to “completely agree.” The main aim of the questionnaires was to help participants shape their thoughts; therefore, all questionnaire data were discarded, except for the general health and demographic information from the introductory questionnaire, open fields with thoughts and questions from the EMAs, and responses to questions on user experience and affinity to technology from the final questionnaire. All questionnaires are available in Multimedia Appendix 1.

Prootyping Phase: The Smart Toilet
Participants made use of an early prototype of a smart toilet seat currently under development at OnePlanet Research Center. The prototype was equipped with electrocardiogram and photoplethysmography sensors, a bioimpedance sensor, a thermometer, and weight sensors. These sensors provide a basic setup that affords the monitoring of the so-called vital signs [44]: blood pressure (electrocardiogram and photoplethysmography), pulse (electrocardiogram and photoplethysmography), body temperature (thermometer), respiration (bioimpedance), blood oxygen (electrocardiogram and photoplethysmography), and weight. Although there are multiple existing methods for measuring the vital signs, what sets the smart toilet apart is its ability to perform measurements automatically and unobtrusively a couple of times a day, which other methods lack, imposing a burden on the user. As stated in the introduction, this basic sensor suite could be expanded to include more sensors that analyze biomarkers in urine and stool and other sensors; however, time and budget constraints necessitate choosing the sensors that would have the most added value. This study was one of the activities performed by OnePlanet Research Center to identify such sensors.

Participants installed the smart toilet seat (see Figure 1 for a schematic overview and Figure 2 for a photograph) with the use of an installation manual, by placing it on top of the regular toilet and fastening the clamps (see Figure 3 for an overview of the installation clamps). They then placed a transmitter device...
within 10 meters of the toilet seat but not necessarily inside the same room; the transmitter needs power from a mains socket, and these sockets are not always available in bathrooms. The transmitter automatically connected to the seat when powered up and sent all the collected data to a cloud-based secure storage. Activation of the sensors on the toilet could be identified through a red light of the photoplethysmography sensor that turned on when the seat came into contact with the skin. The connection of the seat with the transmitter box could be checked through a blue light on the transmitter. Before participants used the seat, one of the researchers checked whether the sensors produced data and whether the data were sent to secure servers.

**Figure 1.** Smart toilet seat prototype. The seat has sensors for measuring physiological parameters and clamps for easy installation.

**Figure 2.** Photograph of the smart toilet seat installed on a regular toilet bowl.
Participants then used the toilet seat for 2 weeks as they normally would, with no additional action needed when visiting the toilet, except from sitting down for urinating, which may be uncommon for some male participants. In the information leaflet, participants could read that the toilet measured their heart rate, body temperature, breathing rate, and weight and that the toilet would transfer this information to a secure cloud storage facility to be able to determine the signal quality. Participants were aware that the data would not be used for any kind of diagnosis or comparison outside of determining the adequacy of signal quality and who was using the toilet; they were also aware that they would not receive feedback on their health or toilet use at any time. After the 2-week period, the seats were disassembled and picked up by the researchers.

Discussion Phase: Focus Group Sessions

In the week after completing the 2-week provotyping phase, all active participants took part in 1 of the 12 web-based focus group discussion sessions, which lasted 45 minutes to 1 hour. These sessions took place through web-based meeting platforms, either Microsoft Teams (Microsoft Corp) or Jitsi Meet (8x8, Inc). The aim was to have 3 to 6 participants and 2 researchers (from a group of 3: the first, second, and third authors of this manuscript) in each session. One researcher played the role of a discussion leader, and the other researcher played the role of an observer and supported the discussion leader where needed.

During the discussion sessions, each participant individually reacted to five discussion theses: (1) their thoughts on the sensitizer materials; (2) their overall experience during the study, such as during the installation, removal, use, and cleaning of the seat; (3) their perceived use cases for the smart toilet seat; (4) how they felt about others knowing about having and using a smart toilet seat; and (5) their opinion on smart toilet seat data privacy. After each participant gave their opinion on a thesis, all other participants had the opportunity to freely react to what they had heard. Every session was recorded; recording was started only after the confirmation of consent from each participant. The session recordings were transcribed and deleted directly after checking the transcription.

To conclude the project, participants received an extensive briefing of the study results, which included the main insights described in this manuscript. The briefing contained no feedback on personal physiological data.

Analysis

Sensitizing and Provotyping Phases

Because the sensitizing questionnaires only served to inform participants’ thoughts about their everyday situation and the use of the smart toilet, we discarded all the data from the sensitizing questionnaires, except for some demographic data (age, gender, and general health status) and the responses to the usability questionnaires (SUS and questionnaires covering hedonic, pragmatic, and efficacy aspects) and the ATI scale. For the user experience questionnaires and ATI scale, descriptive results were calculated: means, medians, and SDs. From the EMA questionnaires, we listed and grouped the open entries with thoughts people had about the smart toilet. No further data from the provocation phase, such as the sensor data, were analyzed in this study. The SUS score was calculated using the following formula:

\[
SUS \text{ score} = \left( \frac{\text{score of items 1} + \text{score of items 3} + \text{score of items 5} + \text{score of items 7} + \text{score of items 9}}{5} \right) + \left( 25 - \frac{\text{score of items 2} + \text{score of items 4} + \text{score of items 6} + \text{score of items 8} + \text{score of items 10}}{5} \right) \times 2.5
\]

which gives a score ranging from 0 to 100. The hedonic quality and pragmatic quality were calculated by taking the mean of the corresponding questionnaire items. The ATI score was calculated by taking the mean of the 9 items and comparing it with the average score of a similar population [43].

https://humanfactors.jmir.org/2023/1/e44850

JMIR Hum Factors 2023 | vol. 10 | e44850 | p.1308
(page number not for citation purposes)
Discussion Phase

Two researchers (MB and SH) manually transcribed the recordings of the discussion sessions. They anonymized the transcript by removing personal information. All transcripts were then read into a qualitative analysis software [45] and analyzed using a method based on inductive thematic analysis [46,47]. Following this approach [47,48], 2 researchers (MB and SH) first performed a primary analysis of 2 session transcripts individually, from which an initial coding scheme emerged, and then compared their coding to ascertain similar interpretations. They further applied inductive coding to identify themes and patterns in the data not yet covered in the coding scheme and then applied the updated coding scheme to the first 5 transcripts. A further iteration of the analysis then took place to ascertain confidence in the coding. The coding scheme was then modified to better reflect emergent themes, and all relevant text segments were coded. This step was repeated until no more issues arose.

Reflexivity

In any research where the researcher attempts to make sense of participants’ experiences, there is a potential risk of researcher bias [49]. To improve the integrity and credibility of qualitative research, researchers must assess how intersubjective components affect data collection and analysis. An instrument for this examination is reflexivity, which refers to researchers’ explicit, self-aware appraisal of their own roles [49,50].

The host institute of the study reported in this manuscript, OnePlanet Research Center, researches potential innovations in health and sustainability using sensor technology and artificial intelligence. One of its research programs is on gut health, in which the smart bathroom is an important part. His work focuses on the human factors research at the center. His work focuses on the acceptability, usability, and efficacy of technological innovations for supporting people in healthy living.

VV is a biomedical field engineer at OnePlanet and is responsible for the design and performance of feasibility and efficacy studies.

MB worked on the research project as partial fulfillment of her Master’s Degree in Science, Management and Innovation at Radboud University Nijmegen.

EW is the principal investigator of the smart bathroom program; she leads all scientific and developmental activities for the smart toilet seat and other innovations.

Results

Participants

Overview

In total, 37 households took part in the study. Of these 37 households, 11 (30%) had >1 active participant, rendering a total of 49 active participants. During the trial, 1 (3%) household containing 1 (2%) participant quit; their data were discarded from the analysis. Of the remaining 48 (98%) active participants, 24 (50%) indicated their sex as male, and 24 (50%) indicated their sex as female. Overall, 28 (58%) participants did not report having any chronic health issues, 2 (4%) participants reported having diabetes, 3 (6%) participants reported having heart problems, 4 (8%) participants reported having asthma or chronic obstructive pulmonary disease, and 8 (17%) participants reported having arthritis. The chronic illnesses mentioned once were bipolar depression, celiac disease, hypertension, heart valve leakage, restless bowel syndrome, and ulcerative colitis.

Participants were, on average, aged 62 (SD 13.97; median 68; range 28-84) years.

Adjustments to Protocol

Unfortunately, we failed to recruit participants from every age bracket as planned. To be precise, only 2 (4%) participants were from the 18-30 years age bracket. However, all other age brackets had >5 (10%) participants as planned. Although the research team put much effort into planning discussion sessions such that they accommodated all participants, it turned out to be impossible to accommodate all the participants because of work schedules, illnesses (especially COVID-19), late cancellations, and the limited availability of participants in the same time frame. Moreover, 2 (4%) participants had to leave the focus group discussion within 10 minutes; they were excluded from the analysis because they did not have the opportunity to contribute to the conversation. Therefore, of the initial 48 active participants from 36 households, 31 (65%) participants from 30 (83%) households completed a focus group session.

Sensitizing and Provotyping Phases

Adjustments to Protocol

Participants were instructed to use the smart toilet seat for 2 weeks continuously; however, 6 (12%) participants were absent for several days and continued to use the seat after returning home, prolonging their provotyping period by the number of days they missed (up to 1 week in 3 cases). One seat needed to be replaced owing to malfunctioning 4 days into the trial; the provotyping period of the corresponding participant was not prolonged. Two (4%) participants filled out their final questionnaire on paper printouts.

User Experience Questionnaires and ATI

The average SUS score was 77.92 (SD 12.81; range 74.20-81.64; 48/48, 100%). This shows acceptable usability and corresponds to the (traditional US) school grading scale of C [39]. To further determine user experience, we also calculated the hedonic quality, which had an average of 4.83 (SD 1.32; range 4.56-5.10; 47/48, 98%) on a scale ranging from 1 to 7. The pragmatic quality was, on average, 4.37 (SD 1.88; range 3.98-4.75; 48/48, 100%). On the ATI, participants scored an average of 3.63 (SD 1.08; range 1.00-5.89; 48/48, 100%), which is similar to a comparison population average of 3.5 [43].

EMA Open Questions

Participants filled out a total of 293 EMAs, with an average of 7.15 (SD 3.95; range 2-18) EMAs per participant. In 165
(56.3%) EMAs, participants did not leave any text in the open remark field, whereas in 128 (43.7%) EMAs, they did, amounting to an average of 3.12 (SD 2.46; range 0-13) EMAs per participant. An overview of the categorization of these remarks is provided in Table 1.

Table 1. Remarks from the open input field of the EMAa questionnaires.

<table>
<thead>
<tr>
<th>Remark</th>
<th>EMA event (n=293), n (%)</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>No remarks entered</td>
<td>165 (65.3)</td>
<td>N/Ab</td>
</tr>
<tr>
<td>Curiosity and compliments</td>
<td>17 (5.8)</td>
<td>“Excited about starting the research”</td>
</tr>
<tr>
<td>Ideas for use cases</td>
<td>15 (5.1)</td>
<td>“Could the toilet detect I tested positive for COVID-19?”</td>
</tr>
<tr>
<td>How does the smart toilet work?</td>
<td>13 (4.4)</td>
<td>“Does the toilet ‘know’ I drank a lot of coffee today?”</td>
</tr>
<tr>
<td>I would like feedback on my data</td>
<td>12 (4.1)</td>
<td>“I would like a display next to the toilet that shows my heart rate and temperature”</td>
</tr>
<tr>
<td>User experience—smart toilet construction</td>
<td>12 (4.1)</td>
<td>“The fastening of the toilet seat is unstable”</td>
</tr>
<tr>
<td>Does the smart toilet work?</td>
<td>8 (2.7)</td>
<td>“I hope the signal comes through; I can’t tell if it works”</td>
</tr>
<tr>
<td>Using the toilet is now automatic or I already forgot about it</td>
<td>8 (2.7)</td>
<td>“I just sit down and don’t think about it anymore”</td>
</tr>
<tr>
<td>Doubts about usefulness; no fun</td>
<td>7 (2.4)</td>
<td>“Will this study produce results?”</td>
</tr>
<tr>
<td>User experience—hygiene</td>
<td>6 (2)</td>
<td>“I find it difficult to clean the toilet with the seat”</td>
</tr>
<tr>
<td>Speaking to or reactions from others about the toilet</td>
<td>6 (2)</td>
<td>“I talked to my son and his friends about the toilet, and they are intrigued”</td>
</tr>
<tr>
<td>User experience—comfort and everyday use</td>
<td>4 (1.4)</td>
<td>“The seat is cold”</td>
</tr>
<tr>
<td>Am I doing this right?</td>
<td>3 (1)</td>
<td>“Do I sit at the right spot? Or not far back enough?”</td>
</tr>
<tr>
<td>Questions about the research</td>
<td>3 (1)</td>
<td>“Does it matter for the measurements that I’m taking medicine?”</td>
</tr>
</tbody>
</table>

aEMA: ecological momentary assessment.
bN/A: not applicable.

Discussion Phase

Adjustments to Protocol

Because of the aforementioned difficulties in planning discussion sessions such that they accommodated all participants, the aim to have focus groups with 4 to 6 people was met for only 19 (61%) of the 31 participants; 10 participants (32%) took part in a focus group with 2 to 3 other participants. The remaining 2 (6%) participants were unable to take part in any of the proposed dates for sessions or expressed a strong preference not to join other participants, which led to these 2 participants being interviewed on their own.

Theme: Everyday Use and User Experience

Approximately half (16/31, 52%) of the participants considered the everyday use of the toilet, including its installation and dismantling when necessary, to be easy and free of complications. A total of 6 (19%) participants mentioned removing the toilet seat for cleaning; 9 (29%) other participants mentioned not removing the toilet seat at all:

- I hardly noticed the difference with my own regular toilet seat. [P623]
- I cleaned it once, no wait, twice, and I removed the seat to do so. Well, that was easy enough. [P280]
- So I just left it on for the whole two weeks; I felt that that would improve the chance that everything would keep working as it should. [P265]

However, most participants (including some who found the everyday use of the toilet uncomplicated) mentioned issues with using the smart toilet seat related to the design of the prototype, the technology, or mismatches with normal practices in using toilets. First, many participants commented on the design of the prototype, with vulnerability, especially of the clamps connecting the seat to the toilet bowl, being the main issue mentioned by 14 (45%) participants:

- If you don’t sit on it correctly, it wobbles a bit. [P283]

Well, using it was not hard, but when you undo the seat, what with all the wires and sensors, if I don’t pay attention and yank too hard, it might well fall off the toilet and everything stops working. So I just left it there. [P265]

A total of 3 (10%) participants felt that the seat, which was a bit higher than a regular toilet seat because of the clamps and weight sensors, made the toilet too high for them:

- I talked to [the researcher] about this, whether my feet could still reach the ground, so we tested that. And it turned out it was way too high, but that wasn’t a problem for the 14 days. [P506]

Furthermore, 8 (26%) participants disliked the color and design of the seat, often citing a mismatch with the overall design of their bathroom:

- Well, the color of the seat, I think brown is a nasty color. It does go well with excrement though. [P850]
I did think that the brown color... My bathroom is all white and blue. I was glad I had my own toilet seat back afterwards. [P580]

Second, the sensor technology in the seat raised questions among some participants. A total of 7 (23%) participants mentioned being intrigued by the red light of the temperature sensor, and for 7 (23%) more participants, this light led to a feeling of being observed:

The red light intrigued me. Sometimes it was on when I got up, and sometimes it wasn’t. Maybe it was not constantly measuring? Or only measuring for 30 seconds? I had no idea. [P166]

It’s not a huge issue, but the red light did trigger a feeling of, well, red means something is wrong. [P781]

Later on, when I had a look at all the sensors, I was wondering which sensor was which, and thought that it could be a camera. I thought that went a bit far, but oh well, it’s all in the game. [P471]

Third, participants reported issues related to a mismatch with everyday habits and practices of toilet use. When the toilet did not match their expectations, this affected their acceptance of the toilet. Hygiene and expectations related to cleaning were the most important issues. All 31 participants mentioned cleaning in one way or another. A total of 15 (18%) participants said that cleaning the seat was difficult because of the clamps, immovability of the seat, and wires. In addition, 14 (45%) other participants mentioned cleaning but also said that they experienced no difficulty, and 2 (6%) participants admitted that they did not clean the toilet at all (and left that task to their spouses):

A normal toilet seat, you can easily put it up and clean the bottom side. This one needs to be taken off entirely. [P515]

I noticed that our cleaning person was anxious to remove the seat, so I told them it was okay to just clean between the clamps for two weeks. But I noticed they were worried about that. [P781]

Also mentioned often (11, 35%, participants) was the fact that the smart toilet seat cannot be put up, which means that all users, regardless of their biology, are expected to urinate sitting down:

I have had some gentlemen visiting, my neighbour and his son. And I tried to talk them into using the toilet as well. But when I said the seat does not go up, they fled! [P400]

As a man, I’m not used to sitting down to urinate, and I found that quite troublesome, especially the first few days. [P768]

Moreover, 2 (6%) participants mentioned that the seat lacked a cover, 1 (3%) participant did not dare use her bidet owing to the fear that it might affect the electronics, and 5 (16%) participants talked about how their “irregular” behavior affected measurements: fidgeting; sitting on the very front of the seat only; sitting on their underpants or bathrobe; and, in 1 case, changing clothes while on the toilet:

I don’t always sit still on the toilet. In the morning, I already take off my pyjama bottoms, and in the evening my trousers, that sort of thing. At one time I started wondering whether that affects the measurements... [P026]

Theme: Perceived Use Cases

A central aim of the study was to find out what use cases potential users would have for the smart toilet. Participants mentioned five categories of use cases (in the order of number of participants mentioning the category): (1) signaling potentially detrimental health conditions or exacerbations of existing conditions; (2) documenting physical data to find out more about oneself; (3) measuring biomarkers to inform a diagnosis; (4) using the smart bathroom for personal science: measuring the results of experiments in lifestyle and nutrition; and (5) tracking biomarkers to inform and trigger behavioral change. No participant mentioned fetishized tracking, that is, tracking out of interest in technology use.

Most participants (25/31, 81%) saw signaling potentially detrimental changes in their health, an early warning system, as a major use case for the toilet seat. This signaling is passive, with measurements occurring in the background. In use cases involving signaling, participants wanted to receive feedback only when there is a need for action:

Someone I know has a heart condition, a leaky valve. She’s ailing a bit but what can you do; I think this would be a solution for her. The seat could notify her in time when her heart condition deteriorates. [P400]

I think the benefit of the seat is that you sit on it regularly, so you get regular measurements and feedback, for instance of blood pressure. That would be important to me. If there’s an outlier, I know I need to do something about it. [P245]

Colon cancer is a real silent killer. Once you have complaints, it’s often too late. If complaints come suddenly, then you notice, but if it comes gradually over a long time, you just don’t realise. And the seat could notice these incremental changes, for instance in how often you need to go. Then you could get a warning that it would be smart to have a colon examination done. [P283]

Overall, 13 (42%) participants mentioned use cases related to signaling critical values associated with their current health conditions, such as cardiovascular disease, diabetes, and gut conditions:

I have ulcerative colitis, which is an inflammation of the gut, and maybe the seat could measure inflammation values in the excrement. And if they are at a high level, the seat could notify me and tell me to make an appointment at the hospital. [P843]

I don’t visit my GP all the time, so it might be that when the GP finds out my blood pressure is too high, it may have been like that for a long period without you knowing. It would be great if I could get a signal that something could be amiss. [P850]
Furthermore, 10 (32%) participants mentioned that signaling could also involve the automatic transfer of relevant information to care professionals:

I would hope that if I would need to see a doctor regularly, that the seat would limit the number of times I have to go there. If it would simply send through the data if something were off, and the doctors could then see that values have changed and we need to act, that would be beneficial. [P471]

I am a cardiovascular patient, even if you cannot tell. And I regularly need to check blood pressure, or fever, or my heart rate. The seat could measure all that for my cardiologist and myself. [P768]

A total of 5 (16%) participants saw uses for the seat as a personal alarm system for older adults living alone:

I have an acquaintance who is seventy years old, and he had a stroke. It was a week before anybody noticed and the police had to break the door. He’s now in the hospital in a serious state. If he had had a smart toilet, the seat could have notified other people that he wasn’t using the toilet anymore...If you’re not going to the toilet anymore, there is something wrong. Maybe you are on holiday, but what if you are just lying there with a stroke? [P515]

Second, participants mentioned documenting their physical state as a major use case for the smart toilet seat: 19 (61%) participants mentioned documenting use cases, that is, registering physiological data to get to know oneself and one’s bodily processes:

I would love to see my own data, to find out what daily rhythms I have. For instance, how long it takes for me to go to the toilet after I have eaten, how fast my digestion is. I would want my data to be available to me to learn such things. [P166]

I would be very interested in sugar content and salt content of my urine. That would give me valuable information about my health. [P781]

Third, 13 (42%) participants mentioned use cases in which the seat can be used for diagnosing health conditions, such as type II diabetes or kidney failure:

I work at a medical laboratory, and we do a lot of urine sampling. The seat would be great to replace a burdensome examination, where people need to collect urine for 24 hours. We could do the first sample in the lab and let the toilet measure the rest...This would be great for diagnosing kidney patients, to check if they produce enough urine. [P026]

What if you could determine if someone has diabetes or prediabetes? If you catch that in time, that could lead to less complications in the future. [P450]

Fourth, 6 (19%) participants would want to use the smart toilet seat for so-called personal science, that is, doing small experiments to find out what affects one’s health:

I have high blood pressure, and I am trying to find a good balance in salt intake. And I would like to know how fast changes in salt intake affect my blood pressure. I just want to check those data every day. [P280]

I stopped eating yoghurt and cruesli after dinner, because my glucose went way up right after that and that affected my sleep quality...I find those trends on a micro-level very interesting. We think we eat healthily but often we don’t always. So I would want to use it for a while to experiment with my diet. [P781]

Fifth, 3 (10%) participants saw use cases involving directive tracking, that is, tracking data to inform behavioral change and habit formation:

Well, I think many people do not drink enough, so waste products cannot leave the body effectively and your urine gets very dark. If the seat could give people feedback on that, and tell them to drink more, that would be good. Also, many people have obstipation, and you can solve that for a great part to eat more fibers and drink more. That’s an easy solution, and if your toilet can tell you that, that would help. [P588]

The toilet seat could measure how much protein a person should ingest. Now, we cannot always measure that, so we use a formula that does not always fit the person. [P450]

The participants mentioned a broad range of conditions that they would like to assess using the smart toilet seat. These use cases ranged from being very vague (“telling me if something is off with my health”) to very specific (assessing salt and glucose levels in urine). A complete overview of the conditions mentioned by the participants is shown in Figure 4.
Other benefits of the smart toilet mentioned by the participants included a relief from burden for themselves and for their health care professionals, cost cutting, better targeted diagnosis, and better care in the period toward or after hospital care:

*I work in the hospital as a dietician. To us, patient weight is very important. Especially older people don’t weigh themselves or have no scales at home. A smart toilet seat could alleviate the work of our nurses because they are under so much pressure, they cannot always weigh our patients. And that limits my work and my advice.* [P450]

*Well, if you need regular check-ups in the hospital, they might arrange matters using the smart seat. Saves you a trip to the hospital.* [P960]

**I have diabetes and would like constant monitoring to replace the finger-prick tests. And while we are at it, cholesterol as well.** [PS80]

Finally, 17 (55%) participants mentioned concerns or doubts regarding the added value or efficacy of the smart toilet. A total of 13 (42%) participants thought that the smart seat had no added value when compared with existing measurement methods. Moreover, 5 (16%) participants mentioned the ways in which feedback on physiological data can have negative consequences, for instance, leading to heightened stress levels. The feasibility of measurements using the smart toilet seat was doubted by 4 (13%) participants, and 1 (3%) participant thought that health practitioners had no capacity to process the data that the smart toilet seat would generate:

*I get nervous when I see a white coat, so my blood pressure rises when I know it gets measured. If I*
would get feedback from the toilet seat, that would probably make me nervous as well. [P506]

What does the toilet have what other devices do not have? You can measure just about anything with a smart watch these days. [P561]

I think that general practitioners aren’t happy when they get all sorts of data that they did not ask for. That will be a very difficult process to manage. [P841]

And you know that if you want to measure blood pressure, you need to sit quietly and not move about. That is not easy on the toilet seat. The moment you sit there, you are already exerting yourself and that is going to influence the measurement. So I don’t know. [P768]

**Theme: Privacy and Data Sharing**

An important theme, mentioned by all participants, was sharing the experience of using the smart toilet seat with others. Participants differed greatly in whether they let others use, or even know about, the seat. A total of 28 (90%) participants talked about experiences of engaging in social interactions regarding the toilet seat (on-stage use), whereas 13 (35%) participants talked about experiences of avoiding social interactions regarding the toilet seat (backstage use). Overall, 17 (55%) participants mentioned only positive sharing experiences; 4 (13%) participants mentioned only avoidance experiences; and 10 (32%) participants mentioned both categories of experiences, embracing social interactions regarding the toilet seat in some situations and avoiding them in others (off-stage use):

> My neighbour and some other visitors, I led them to my upstairs bathroom, and told them ‘Have a look, will you?’ Another friend, I wanted to show her the smart seat, but she thought it would not suit her husband. If she were alone, she would want to try it at home as well. And I asked her whether she wanted to see the seat, but she didn’t. [P400]

Well, it always led to conversations, right? Especially if there were young people who needed to use the toilet. I did warn them in advance, told them not to be scared. But everyone thought it was interesting and had all kinds of questions. I just told them it was something new, and maybe they would have something to do with it in the future. [P040]

> It’s nobody’s business. We don’t mind, but you don’t need to discuss toilet seats with your guests. Not because of etiquette or anything, but you just don’t. [P214]

The participants who interacted with others regarding the seat did so on different levels. Overall, 16 (52%) participants mentioned talking about the smart seat with others, 6 (19%) participants mentioned showing the smart seat to visitors, and 10 (32%) participants mentioned letting other people use the smart seat. A total of 15 (48%) participants mentioned receiving positive responses when interacting with others about the smart seat, 4 (13%) participants mentioned receiving negative responses from others, and 4 (13%) participants mentioned receiving both positive and negative responses. With regard to positive responses, 13 (42%) participants mentioned others showing interest, 7 (23%) mentioned others having questions, 3 (10%) participants mentioned others being surprised, 3 (10%) participants mentioned others showing acceptance, and 1 (3%) participant mentioned that a visitor wanted to take part in the study as well and try the seat at home:

> I talked quite a bit about it while walking the dog. I run in to a lot of dog owners and we chat, and then I would talk about the seat. People are very interested; they like to hear about it. [P471]

> It does evoke questions, you know. That makes sense, because suddenly there is an extra seat with a red light. So I can imagine people wonder. But that was not an issue, it just took some explaining with some people. [P500]

Participants who did not interact with others regarding the smart seat had different strategies to avoid interaction. A total of 11 (35%) participants mentioned having toilets on different floors in their house and installing the smart seat in their upstairs bathroom so that visitors could use the downstairs toilet and not see the smart seat, 4 (13%) participants explicitly expressed not mentioning the smart seat to others, 2 (6%) participants took off the smart seat whenever people visited, and 1 (3%) participant mentioned the fact that nobody visited them during the 2-week period:

> I did that on purpose [installing the toilet in the upstairs bathroom]. I reckoned that if I have visitors, I don’t want those people on that seat. That will just give rise to questions and remarks. So I just skirted around the issue; I thought let’s not have that. [P623]

> We have the privilege of having two downstairs toilets. I just used the one, and my wife used the other. But we did think about what the consequences would be if we did not have that; if we had visitors, we would have had to take the seat off and explain all kinds of things, and we would not want that. [P721]

When it comes to sharing their data, all but 2 (6%) participants had reservations. A total of 15 (48%) participants explicitly mentioned that they want ownership of their data so that they can control who can see or use the data; 4 (13%) participants would only want to share secondary, derived data, such as blood pressure averages and trends, and not the raw data; 5 (16%) participants would share the data but only when anonymity can be guaranteed; and 12 (39%) participants said that they are suspicious or worried about data sharing:

> What is important to me is that I have the say over the data. I want to decide who I share data with. Whether that’s my GP, or my neighbour to compare our physiology. [P166]

> Personally, I think online privacy is a false sense of security, especially with smartphones. Every click you make gets measured by algorithms and sold to third parties. [P561]
I am a bit scared that insurance companies and the like will draw all kinds of conclusions from the data. [P280]

Among the 22 participants who would share their data, all but 2 (9%) participants would limit accessibility. A total of 18 (82%) participants would share their data with medical professionals, 7 (32%) participants would share their data with scientists, and 2 (9%) participants would share their data with the producer of the product. A total of 7 (32%) participants explicitly mentioned not wanting to share their data with the producer of the product, 4 (18%) participants explicitly mentioned not wanting to share their data with the government, and 7 (32%) participants explicitly mentioned not wanting to share their data with the commercial sector:

My limit is the doctor who needs the data to help me. [P580]

I can imagine health care professionals getting the data, that would not be so strange, but I still think I need to actively consent to transfer the data to them. [P588]

If there is one group that I don’t trust it’s the government, unfortunately. If you look at the recent scandals...And they have computer systems that don’t work all that well... [P484]

Barriers to and Requirements for the Use of Smart Toilets in the Home

Textbox 1 display an overview of the barriers to and requirements for the use of smart toilets in homes.

<table>
<thead>
<tr>
<th>Usability and everyday use</th>
</tr>
</thead>
<tbody>
<tr>
<td>• The smart toilet should fit the current design of regular toilets.</td>
</tr>
<tr>
<td>• The smart toilet should be sturdy and not be easily breakable.</td>
</tr>
<tr>
<td>• The toilet should be adjustable in height, as it is currently too high for some users</td>
</tr>
<tr>
<td>• The smart toilet should match the color and design of the regular toilet.</td>
</tr>
<tr>
<td>• The smart toilet should afford toilet habits, such as putting the seat up or closing the cover lid.</td>
</tr>
<tr>
<td>• The smart toilet should be easy to clean.</td>
</tr>
<tr>
<td>• The smart toilet should be inconspicuous so that privacy of use is warranted if desired.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data agency</th>
</tr>
</thead>
<tbody>
<tr>
<td>• The smart toilet should provide users with not only full access to their data but also the option to receive feedback only if there is a need for action or grave concern (signaling).</td>
</tr>
<tr>
<td>• The smart toilet should provide understandable and actionable feedback on relevant biomarkers and health data.</td>
</tr>
<tr>
<td>• The smart toilet should afford the option to share data or derived data with carers, general practitioners, or other medical professionals.</td>
</tr>
<tr>
<td>• The smart toilet should afford complete user control over further data sharing.</td>
</tr>
</tbody>
</table>

Sensitizers

Overall, 16 (52%) participants made remarks about the sensitizing questionnaires. A total of 11 (35%) participants mentioned the positive aspects of the questionnaires, mostly about the ease of use, whereas 14 (45%) participants mentioned negative aspects, mainly about questions they thought were irrelevant, such as those on the beauty of the prototype, or hard to answer, such as questions on mood and stress:

I liked the questionnaires, and they made me think about my role in the research. I started wondering about my data, and what [the researchers] would use them for. I did receive instructions beforehand, of course, that it would be about my experience and what using the smart toilet evoked in me, and those questions surely augmented that. [P843]

Well, I thought the questionnaires were a bit dodgy. To me, the seat was the seat and nothing else, just like any seat. So I did not see the added value of the questions. There’s no feedback, so if I sit on my own toilet seat or yours, it’s all the same to me. [P850]

There was a question about whether I thought the seat was beautiful, and I thought that made no sense. What is beautiful about a toilet seat? A white one or a brown one, it does not make a difference on how I sit. Well, I don’t really like the brown color, but it’s all part of the game. [P721]

Saturation

Saturation was determined by calculating the number of unique themes for a base run of 4 transcripts and then establishing the percentage of new information coming forth from adding additional runs of 3 transcripts [51]. Top-level theme saturation was reached in the base run, with new top-level themes emerging below the 5% threshold for each additional run. Code saturation was achieved after including the first additional run, with new codes emerging below the 5% threshold for each additional run.

Discussion

Principal Findings

This study aimed to investigate how people experience using a smart toilet installed in their home: their perceived use cases
for the innovation, the limitations and barriers to its everyday use, sharing the experiences with the toilet seat with others, and privacy and data sharing concerns. The results revealed that participants already found the current prototype quite usable, but most participants mentioned issues that can not only inform future iterations of the prototype but can also elucidate people’s expectations of smart bathroom technology. These expectations had a strong association with norms and behaviors around toilet use. The fact that the seat could not be raised, which entailed being seated when using the toilet, was problematic for many of the male participants and male visitors to the participating households. The fact that the current prototype was difficult to clean, especially because of the way it was connected to the toilet bowl, was mentioned by almost all participants, except for the 2 (4%) participants who admitted leaving toilet cleaning to their spouses. The fact that even for a prototype, the color and form play a role in acceptance shows that these aspects need to be considered when developing future iterations. On a more general level, this result shows that smart appliances need to fit everyday practices and norms.

Participants provided a broad range of use cases for the smart toilet seat. They saw signaling undetected health conditions or exacerbations of existing conditions as the most important potential application. Signaling occurs in the background, without notification or feedback, unless a result that warrants attention pops up. Further much-mentioned use cases were documenting all kinds of physiological and behavioral data to better understand oneself and using the toilet seat to diagnose certain conditions (which differs from signaling in that it is an active, overt process). Other use cases were personal science, in which the toilet seat is used to measure the effects of experiments with nutrition on participants’ health and using the toilet seat as a driver of behavioral change. These differ from existing frameworks in the literature [29,30], which lack signaling medical conditions but do cover use because of interest in the technology. The difference between these frameworks and the current results lies in the research sample. The cited studies included people interested in lived informatics and quantifying self-movement. Such people would be likely to actively adopt trackers, for instance, to measure their physical activity or heart rate variability. This study included a broader range of participants with a broader range of interests in technology per se and in the measurement of their own data. This broader range of interests is expressed in the number of participants who expressed worries about how feedback on physiological data may raise their stress levels or who do not see the added value in the smart seat. Further research can shed light on whether this sample better reflects the attitudes in the general population than the frameworks from lived informatics research.

The issue of data agency is a recurrent theme throughout the results of this study. When talking about their perceived use cases for the smart toilet seat, many participants expressed a desire for acting with and upon and learning from their data (eg, personal science use cases), whereas others expressed an opposite desire, that is, for data to be hidden from them unless there is something important that they need to act upon right away. This indeed shows that people must have the autonomy to determine the level of data availability by themselves for the technology to fit their needs. When talking about data ownership and privacy needs, the importance of data agency becomes even clearer. Data ownership and privacy protection are needs that must be met.

The many concerns participants expressed around sharing their experiences and their privacy needs confirm earlier research [31,52] and show that these issues should play a larger role in the development of smart home appliances. The study confirms work that shows that people have different needs when it comes to the on-stage, off-stage, or backstage use of technological innovations. Some participants were willing to present the smart seat to visitors and even go so far as to invite people into their homes to do so. Some were more reluctant and would discuss the smart seat only when the need to do so arose, and others, the backstage users, avoided sharing their experiences altogether, for instance, by “hiding” the smart seat in an upstairs bathroom. This study also shows that the same people can show different presentation preferences toward different people; what one shares with a close friend may differ from what one shares with a neighbor. The fact that this pattern is already present in a study with self-selection of participants (see subsequent paragraphs for the discussion of self-selection bias), in which we can expect more people who have no qualms about using or discussing toilets to participate, may very well mean that it is even more pronounced in the general population because people who avoid discussing this topic altogether could be less likely to take part in this research. This has consequences for the acceptability and design of smart appliances that are integrated into the home: it should be possible to put them away or hide them or their design should be inconspicuous.

Privacy and ownership of data in smart home appliances for health have been the focus of attention for at least a decade (eg, the study by Townsend et al [53]). The participants showed a strong preference for the ownership of their own data and having responsibility for sharing, transparency in who uses their data and for what purposes their data are used, and protection from undesired consequences. This reflects the findings of many other studies (eg, the studies by Kennedy et al [54], Forchuk et al [55], and Choi et al [56]). However, these concerns, as yet, have not been taken into account when developing actual products that enter the marketplace; very few of these products make the user the owner of their own data or provide them with the opportunity to control the flow of data and access. In future innovations, data management and privacy should play a more important role.

The study shows that the approach we followed, which consisted of sensitizing, provotyping, and discussion, was a successful method for supporting participants to voice their thoughts and concerns. The sensitizing phase succeeded in making people think about the smart toilet and various health subjects before the trial began. However, the participants’ responses also showed that sensitizing materials must be carefully designed. In this study, some participants showed irritation or other negative reactions because of questions they did not see the point of, such as the questions on esthetic aspects, part of the user experience questionnaire covering hedonic aspects. Some participants did not see the relevance of answering questions
about the “beauty” of a prototype, as it was obviously not the finished product. Interestingly, their irritation did make them consider and talk about esthetics, a facet of the prototype design that they would otherwise never have thought of. However, to ensure that participants are not alienated by the sensitizing materials, these materials should be better pilot-tested and more carefully worded. Moreover, the burden of the sensitizing phase should be as low as possible.

In this study, prototyping proved very successful. First and foremost, it gave participants the necessary experiences for talking about barriers and needs surrounding smart appliances used in sensitive areas of the home. Moreover, the approach succeeded in making the normally unsaid factors available for discussion: norms, taboos, and cultural practices that are so embedded in everyday life that they escape conscious scrutiny. The most important example in this study were the conversations on participants’ toilet use habits, which they would normally never talk about. These conversations presented valuable insights that may even go beyond the current research setting: there is surprisingly little, if any, literature on the everyday practices of toilet use; the current literature only mentions toilet use when practices are greatly different from the Western norms and standards, such as works on communal toilet facilities in South Africa [57] and East and Southeast Asia [58,59] or on latrine use in rural India [60]. A second source of literature on toilet use is a side note in a work on ensuring sustainability through water use reduction; here, toilet use is mentioned as being “highly routinised” and therefore “very difficult to change” [61,62]. The current results are, therefore, also interesting as an ethnography of toilet use practices, especially when it comes to aspects of toilet use that are so embedded in everyday life that they usually remain unsaid in scientific discourse: standing up while urinating, lifting and lowering of toilet seats, and hygiene aspects.

The discussion phase served its purpose and delivered a rich qualitative data set. Unfortunately, it proved impossible to have all participants join focus groups of 3 to 4 people to obtain the desired group size. The number of participants who ended their prototyping phase in roughly the same time frame was limited by the number of toilet seats we had at our disposal; moreover, the COVID-19 pandemic and limited availability of many participants also played a role. Taking part in smaller groups means that although every participant gets ample time to share their thoughts, they have less opportunity to get inspired by what others say and hear different voices. This limitation may have reduced the richness and value of the data in this research.

Although we did our best to eliminate bias, no research can escape potential influences on validity. The sample we included in this study is likely to have a certain amount of participant bias because of self-selection, which is, for example, visible in the age of the participants. Even though we aimed to include people from all age groups older than 18 years in similar numbers, we did not manage to achieve that; the average age of the participants was 62 (SD 13.97; median 68; range 28-84) years, and more than half (25/48, 52%) of the participants were aged between 60 and 84 years. This may have affected the results because the results now mostly reflect the viewpoints of older people interested in taking part in this kind of research. However, the age of the participants could also be seen as an indication of potential interest, with older users being naturally more inclined to be concerned about their health in general and their gut health in particular. Furthermore, the ATI scores of the participant group resemble those of a comparable general population, which means that the participants are likely to be representative of a broader audience when it comes to interest in technology use in everyday life. Finally, the self-selection bias could mean that the reservations participants mentioned about sharing experiences regarding the smart toilet and about data ownership and privacy could very well be more pronounced in the general population, as people who have very strong reservations are unlikely to take part in this kind of research.

A self-serving researcher bias may have arisen from the aims of the research program. Members of the research team were deeply involved in the development of the smart toilet, which may have curbed the participants’ inclination to express negative opinions about the seat. However, the main interviewers (SH and MB) had no such vested interests in the success of this prototype; moreover, the results show that the participants felt free to cast their doubts, saying that they have no use for the smart toilet or feel skeptical about its efficacy.

Third, as stated before, because the current scientific literature on toilet habits and use is lacking, it is difficult to evaluate the generalizability of the results of this research. Owing to logistic limitations, all participants came from the province of Gelderland in the Netherlands and its neighboring regions, so the results found in this study might theoretically be limited to this region. However, when it comes to the current toilet design and use, this region can be seen as representative of large swathes of the global population. In the European Union, 98% of the population has similar toilets [63], and so does the US population [64]. The results from this study almost certainly would not apply to many people in Asia, for instance, those in China and India, where squat toilets are ubiquitous. Further research can elucidate whether our hypothesis that the current results are valid for those parts of the world that have similar-style toilets is correct.

The smart toilet described in this research is not unique. There are a number of similar initiatives around the world, both in academia and industry, such as the Stanford smart toilet [65], Toto smart toilet [66], and Rochester Institute of Technology smart toilet [67], and there are modules that can be placed inside regular toilets to measure urine contents, such as the Withings U-Scan (see the description in the study by Sequeira-Antunes and Ferreira [68]). However, research on and toward these toilets and modules has as yet concentrated only on technical efficacy. To our knowledge, there has not been any research on use cases as seen by potential users; the barriers to and facilitators of acceptance; and other issues of use in general daily life, such as fit with culture and habits. Our study not only sheds light on our own prototype in these regards but can also inform the design and development of other endeavors in the field.

Finally, this study concentrated on the use cases, needs, and barriers put forward by potential end users. However, the acceptance and efficacy of smart innovations in health care depend on many more stakeholders, including health care
professionals, social workers, health insurance providers, and public policy makers. Further research will include their voices as well.

Conclusions

This study showed that participants felt that a smart toilet seat could be acceptable and effective, as long as it fits everyday practices concerning toilet use and hygiene. The use cases they envisioned ranged from signaling the deterioration of health conditions to documenting health data to informing diagnoses to engaging in personal science endeavors to driving behavioral change. Participants differed in how much they wanted to share their use of the smart toilet with others; whereas the majority (17/31, 55%) shared their experiences of using the toilet with others, 4 (13%) participants never talked about the toilet with others or let others see or use the toilet, and 10 (32%) participants shared their experiences with some people but not with others. When it comes to the data produced by the smart toilet seat, participants expressed a need for ownership, transparency, and control; most participants (18/31, 58%), however, would share their data with health care professionals. Finally, the method used in this study proved to be a successful way to support people in talking about aspects of their behavior and everyday life that normally remain unspoken. The results of the study not only inform further iterations of the smart toilet prototype and the smart bathroom program but also have relevance outside these applications. The categories of use cases mentioned by the participants differ from those in the current literature and may provide a better reflection of average users than the categories of use cases mentioned in studies from the realm of quantified self-movement. Using or avoiding the use of technology for self-presentation is a relatively underresearched topic, which may, however, have a great impact on the acceptance and public use of smart appliances, wearable technology, and other technologies for supporting people’s health. Future research on these subjects can further strengthen our knowledge.

Acknowledgments

OnePlanet Research Center acknowledges the province of Gelderland for their funding.

Data Availability

The data sets used or analyzed during this study are available from the corresponding author upon reasonable request.

Authors’ Contributions

SH, VV, MB, and EW conceived and designed this study. SH, VV, and MB led the discussion groups. All authors conceived all analyses, and SH and MB performed them. SH wrote the first draft of the paper. All the authors reviewed the paper, made key intellectual contributions to the content and reporting, and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Sensitizing questionnaires.

ROCX File, 23 KB - humanfactors_v10i1e44850_app1.docx

References


64. Residential toilets. United States Environmental Protection Act. URL: https://www.epa.gov/watersense/residential-toilets [accessed 2022-12-01]


Abbreviations

ATI: Affinity for Technology Interaction
EMA: ecologic momentary assessment
SUS: Systemic Usability Scale

©Sander Hermsen, Vera Verbiest, Marije Buijs, Eva Wentink. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 11.08.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Extending the Privacy Calculus to the mHealth Domain: Survey Study on the Intention to Use mHealth Apps in Germany

Abstract

Background: With the increasing digitalization of the health sector, more and more mobile health (mHealth) apps are coming to the market to continuously collect and process sensitive health data for the benefit of patients and providers. These technologies open up new opportunities to make the health care system more efficient and save costs but also pose potential threats such as loss of data or finances.

Objective: This study aims to present an empirical review and adaptation of the extended privacy calculus model to the mHealth domain and to understand what factors influence the intended usage of mHealth technologies.

Methods: A survey study was conducted to empirically validate our model, using a case vignette as cover story. Data were collected from 250 German participants and analyzed using a covariance-based structural equation model.

Results: The model explains R²=79.3% of the variance in intention to use. The 3 main factors (social norms, attitude to privacy, and perceived control over personal data) influenced the intention to use mHealth apps, albeit partially indirectly. The intention to use mHealth apps is driven by the perceived benefits of the technology, trust in the provider, and social norms. Privacy concerns have no bearing on the intention to use. The attitude to privacy has a large inhibiting effect on perceived benefits, as well as on trust in the provider. Perceived control over personal data clearly dispels privacy concerns and supports the relationship of trust between the user and the provider.

Conclusions: Based on the privacy calculus, our domain-specific model explains the intention to use mHealth apps better than previous, more general models. The findings allow health care providers to improve their products and to increase usage by targeting specific user groups.

Keywords

mHealth; mobile health; confidential; privacy calculus; privacy; intention to use; adoption; data autonomy; social norms; trust in the provider; trust; privacy concern; benefit; attitude to privacy; survey; intention

Introduction

Background

The use of digital health products, which promise to increase the effectiveness and efficiency of health care delivery, is on the rise. Between autumn 2019 and summer 2021, downloads of mobile health (mHealth) apps in Germany doubled to 2.4 million [1]. mHealth apps run on mobile devices and may provide medical services ranging from individual care to public health measures [2]. They are said to improve individual health competence and, ultimately, motivate users to deal with their own health more responsibly through interventions and access to information, simplified communication with experts, and the tracking of health data [3-5]. In addition to these advantages, there are also risks associated with using mHealth apps. For example, the security infrastructure of many apps is currently inadequate and does not meet the requirements for protecting
user data (eg, the General Data Protection Regulation [GDPR] in the European Union and the Health Insurance Portability and Accountability Act [HIPAA] in the United States) [6]. It is therefore not surprising that mHealth users are becoming increasingly sensitive to data privacy and data security [7-9]. Given the pros and cons of using mHealth technologies, it is essential to take a close look at the factors that influence users’ intention to (not) use them in order to inform and improve mHealth technology design and, ultimately, increase the uptake of safe and efficient technologies. To examine why people intend (not) to use mHealth apps, we decided to build on the privacy calculus model.

In this study, we focus on the use of health insurance apps because, on the one hand, there is already a large number of users and, on the other hand, a large number of potential users due to the mandatory membership in a health insurance company in Germany [1].

Related Work
The privacy calculus model originally postulated that users of social network sites (SNSs) perform a calculus between the expected loss of privacy and the potential gain of disclosure when deciding whether to use it [10]. That is, the model suggests that people compare potential benefits and costs to calculate their intention to use the SNS technology [11-13]. If the sum of the drivers (benefits) is greater than that of the inhibitors (costs), people will use the technology. If the number of inhibitors is greater, the use of the technology is rejected [11,14,15]. The privacy calculus model was successfully used to predict the intention to use SNSs [16] and e-commerce websites [17]. Based on the privacy calculus model, we aim to understand which factors have a concrete influence on the cost-benefit calculation underlying the intention to use mHealth apps.

Thus far, 3 studies that have examined the intention to use mHealth apps based on the privacy calculus model. They were limited either by the lack of explained variance ($R^2$ values did not exceed 0.5 [11,18] or were not reported [19]) or marginal model fit values [20], which indicate that the used model did not properly fit the observed data [19]. Conceptually, we think these studies [11,18,19] underrepresented the following 3 domain-specific factors influencing the intention to use mHealth technologies:

- When examining the intention to use mHealth technology, the data autonomy granted to the users, that is, the control over granular privacy settings to limit access to their data [14,16], was not taken into account [18,19] or only partially accounted for via the concepts of privacy concerns [11]. Studies have shown, however, that data autonomy influences the intention to use data-collecting mHealth technology [21].
- Although the direct or indirect influence of trust in the provider on the intention to use mHealth technology has been examined in 2 studies [11,19], the individual’s interest in the object represented in the trusting relationship—here the protection of personal data—has not been considered [22]. If the user is not interested in the security of personal data, a relationship of trust concerning the use of data would be irrelevant. Consequently, to be able to make statements about a trusting relationship, the general attitude to privacy should be considered [22,23].
- None of the existing studies considered the influence of social norms, such as social pressure from family and friends. However, there is evidence that social norms influence the acceptance of mHealth technology for disease prevention, especially in healthy individuals [24,25].

Aim of This Study
To achieve our overall goal (ie, to explain the intention to use data-collecting mHealth technology), we address 3 subgoals in this article: (1) we investigate whether perceived data autonomy reduces privacy concerns and has a positive effect on the intention to use mHealth apps, (2) we explore the influence of an attitude to privacy on trust in the provider, and (3) we examine the influence of social norms on the intention to use mHealth apps. To implement these subgoals, we first explain our model and derive hypotheses. We then validate our model in a survey study using a covariance-based structural equation model (CB-SEM). After discussing the results, we derive theoretical and practical implications and reflect on the limitations of the study. We end our paper with a conclusion concerning our objectives.

Model Description and Hypotheses
To predict and examine the intention to use mHealth apps, we adapted a privacy calculus model from the SNS domain [12]. In contrast to privacy calculus models in the mHealth area, in the SNS domain it is common to examine the influence of social norms and perceived data autonomy. Therefore, in addition to the constructs of perceived benefits, privacy concerns, and trust in the provider, the adapted model also included the constructs of perceived control over personal data (subgoal 1) and social norms (subgoal 3) [12]. Finally, we added the attitude to privacy to the model to cover subgoal 2 from above. Unlike previous studies [11,18,19], we refrained from adding health-specific factors (eg, health concerns) to reduce the complexity and increase general applicability of the model. Figure 1 shows the final model with drivers (+) and inhibitors (−), which we will elaborate on in turn.
**Perceived Benefits**

Perceived benefits are both the hedonistic and the utilitarian reasons people may have to use a product or service. Hedonistic reasons may be that the process of using a technology is fun and enjoyable, irrespective of what may be achieved by using it [16,26]. On the other hand, utilitarian reasons are mainly associated with an increase in productivity and efficiency (e.g., time savings, economic advantages) [17,27,28]. In the area of mHealth, utilitarian advantages may also relate to the simplification of treatments and coordination between different medical institutions, which can lead to more efficient treatments and, ultimately, better health outcomes [4,5,11]. There is evidence that the perception of benefits has a driving influence on the intention to use data-collecting and disclosing mHealth information technology [4,21].

- **H1**: Perceived benefits positively influence users’ intention to use mHealth apps.

**Privacy Concerns**

Privacy concerns describe users’ concerns about a possible loss of privacy using web-based apps due to privacy risks, such as data leaks and data misuse [15]. These concerns are driven by situational risk perceptions, for example, data that are not secure with a particular provider [15]. Thus, privacy concerns can be thought of as a situational motivator to be careful when disclosing personal data [14,29,30], and, ultimately, to inhibit the use of health technologies that require disclosure of personal data [21,31,32].

- **H2**: Privacy concerns negatively influence users’ intention to use mHealth apps.

**Trust in the Provider**

Trust is a complexity-reducing variable because it makes the trustor bear a perceived risk when cooperating with a trustee [33]. In other words, trust is a psychological state where a person accepts being vulnerable to the actions of another party because the person expects that the other party will carry out a certain action in their interest, regardless of whether the action is monitored [34,35]. When interacting with information technology, people’s focus is less on trust in the functionality of the system and more on trust in the provider to protect their data and privacy [36,37]. Various studies have shown that trust in the provider has a significant positive influence on the acceptance of mHealth technologies and their intended use [3,38-41].

- **H3**: Trust in the provider positively influences users’ intention to use mHealth apps.

**Social Norms**

Social norms are social and psychological factors that are inherent in group dynamics and strongly influence individual human behavior [14]. People tend to behave in ways that are (socially) accepted to continue to benefit from the advantages of being part of a social group (injunctive norms). Violation tends to be punished with disapproval and possibly social ostracism [14,42,43]. Besides, individuals follow the behaviors of others (descriptive norms) [43]. In the case of health prevention through mHealth technology, users’ intention to use mHealth technology is influenced by both the approval of technology use in their social environment (e.g., injunctive norms friends and family) and the descriptive norms based on how and when a technology is used in the social environment [12,24,44].

- **H4**: Social norms positively influence users’ intention to use mHealth apps.

**Perceived Control Over Personal Data**

Perceived control is a psychological construct that describes individuals’ perceptions of the extent to which they can influence and control the achievement of a certain goal and the resources that are necessary to do so [11,45]. In the context of mHealth apps, this involves the perceived ability to control which health data are collected and who can access them.
Various studies have shown that if control over personal data is perceived to be limited, privacy concerns increase [8,11,46]. By contrast, if people think that they can control their data, their intention to use mHealth technology [8,11] and their trust in the technology provider increases [11,22,33,40].

- H5a: Perceived control over personal data positively influences users’ intention to use mHealth apps.
- H5b: Perceived control over personal data negatively influences users’ privacy concerns.
- H5c: Perceived control over personal data positively influences users’ trust in the provider.

**Attitude to Privacy**
We define the attitude to privacy as a user’s general tendency to consider privacy and data security to be important or a user’s disposition to value privacy [15]. The inclusion of this construct in the privacy calculus model is particularly important in the mHealth context because disclosure of health data tends to be more consequential than data stored on other technologies, such as SNSs [47]. A strong attitude toward data protection has an inhibiting effect on people’s intention to disclose data (ie, their privacy concerns) and their intention to use a data-collecting technology altogether [15,28,48]. Once data have been disclosed, users with a strong attitude to privacy are more interested in the whereabouts of their data and consequently more cautious when it comes to trusting the provider using their private data [22,49-51]. Finally, whereas the perception of potential risks may be overinflated due to strong attitudes to privacy, potential benefits of technology use may be undervalued [52-56].

- H6a: Attitude to privacy negatively influences users’ intention to use mHealth apps.
- H6b: Attitude to privacy positively influences users’ privacy concerns.
- H6c: Attitude to privacy negatively influences users’ trust in the provider.
- H6d: Attitude to privacy negatively influences users’ perceived benefits.

Now that we have explained the theoretical basis of our model, we evaluate the underlying hypotheses in a survey study. In the next section, we describe the methodological basis of this study.

**Methods**

**Participants**
The theoretical framework described in Figure 1 was empirically tested using data gathered via an online survey that was performed as part of a bigger study in cooperation with a German health insurance company (BARMER), one of the largest and best-known health insurance companies in Germany. The survey was administered by a commercial survey agency in Germany (Norstat GmbH), which also organized the entire survey process (programming the online study and collecting the data). We targeted a sample of at least 250 participants to be able to calculate the model validly [20]. Participants were individuals who registered with Norstat GmbH as survey participants. In addition to being a resident of Germany and a native German speaker, the prerequisites were that the participants were customers of a German health insurance company, as the case vignette centered on a German health insurance app. The minimum age for participation was 18 years, as this is also the minimum age for admission as a Norstat panel member. There were no prerequisites regarding gender. Data collection took place from March 11, 2021, to March 17, 2021. Our estimated minimum time to complete the survey was 5 minutes. This was ensured by the system allowing participants to continue the survey only after a certain amount of time (60 seconds for the consent form, 30 seconds for the case vignette, and 210 seconds for the questionnaire). The mean and median participation times were both 6 minutes with a standard variation of 42 seconds. Participants volunteered to participate after giving informed consent and received compensation ($0.80 [US $0.90]) for taking the survey.

**Ethical Considerations**
Because a third party (Norstat GmbH) contacted potential participants and collected the data, we did not have direct contact with participants or access to any personally identifying participant information. We obtained only completely anonymous data. Consequently, we were able to guarantee full anonymity and privacy of the participants, which conforms to the ethical guidelines of the German Research Foundation. Thus, based on the guidelines of the Ethics Committee of our Institute (Institute of Psychology and Ergonomics) no additional ethics board review was mandatory [57].

**Materials**
Following a practice that is often used in technology acceptance studies [58], the study used a case vignette to evoke a typical situation where an mHealth app would be used and described the trade-off between the benefits of using it and its data privacy risks. We decided to describe a health insurance app in the case vignette because, as already described, they currently account for the largest share of mHealth app downloads in Germany [1]. In particular, the case vignette (Multimedia Appendix 1) describes a situation in which a friend “Alex” uses the app of his health insurance on a wearable to track his health behavior (ie, physical activity). By participating in the bonus program of this insurance, Alex may receive a bonus of up to €100 (US $112) for working out regularly (a direct benefit), but the insurance may also deny covering treatment costs due to an unhealthy lifestyle (a possible risk). To assess the factors included in the privacy calculus model displayed in Figure 1, we used a 30-item questionnaire (Multimedia Appendix 2; also see [14,16,27,42,53,59-62]), which we developed following the methodology of Moore and Benbasat [63]. All items were measured on a 7-point Likert scale that ranged from 1 (strongly disagree) to 7 (strongly agree).

**Procedure**
The survey consisted of 3 parts. In the first part, demographic data of the respondents were recorded, such as age, gender, and educational level. In the second part, the respondents were asked about their individual experience with mHealth apps as well as their current use of wearables such as fitness trackers and smartwatches (also beyond health apps). In the third part, the participants received the case vignette and were asked to answer questions about their individual experience with mHealth apps as well as their current use of wearables such as fitness trackers and smartwatches (also beyond health apps).
the questionnaire. The order of the questions in the questionnaire was randomized for each participant.

Analyses

To test the model outlined in Figure 1, a CB-SEM was used, which is a common approach to theory testing and confirmation [64]. The CB-SEM was carried out with lavaan [65] (version 0.6-9; R Foundation) in RStudio (version 1.3.1093; Posit, PBC), using the maximum likelihood estimator. All items of the questionnaire were included in the analysis and restricted to load on the respective constructs described above and in Figure 1.

Results

Survey Characteristics

A total of 336 observations were collected. After deleting observations that were unusable because of missing responses, a final sample of 250 observations (126 male and 124 female) was used for further analysis. The mean age of participants was 46.5 years (SD 15.2 years). The demographic characteristics of the sample are summarized in Table 1.

Table 1. Demographic data of the sample (N=250).

<table>
<thead>
<tr>
<th>Demographic characteristic</th>
<th>Frequency, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>126 (50.4)</td>
</tr>
<tr>
<td>Female</td>
<td>124 (49.6)</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
</tr>
<tr>
<td>No degree</td>
<td>2 (0.8)</td>
</tr>
<tr>
<td>School leaving certificate</td>
<td>39 (15.6)</td>
</tr>
<tr>
<td>Secondary school certificate</td>
<td>88 (35.2)</td>
</tr>
<tr>
<td>General qualification for university entrance</td>
<td>57 (22.8)</td>
</tr>
<tr>
<td>University degree (bachelor’s or master’s)</td>
<td>62 (24.8)</td>
</tr>
<tr>
<td>PhD</td>
<td>1 (0.4)</td>
</tr>
<tr>
<td>Other</td>
<td>1 (0.4)</td>
</tr>
<tr>
<td><strong>Experience with mHealth(^a) apps</strong></td>
<td></td>
</tr>
<tr>
<td>Regular use of mHealth apps</td>
<td>124 (49.6)</td>
</tr>
<tr>
<td>Occasional use of mHealth apps</td>
<td>34 (13.6)</td>
</tr>
<tr>
<td>No use of mHealth apps</td>
<td>92 (36.8)</td>
</tr>
<tr>
<td><strong>Usage of wearables</strong></td>
<td></td>
</tr>
<tr>
<td>Regular use of wearables</td>
<td>73 (29.2)</td>
</tr>
<tr>
<td>No use of wearables</td>
<td>177 (70.8)</td>
</tr>
</tbody>
</table>

\(^a\)mHealth: mobile health.

Assessment of the Structural Model

The internal consistency of the scales as well as convergent validity and discriminant validity of the measured constructs are shown in Tables 2 and 3. Internal consistency was evaluated with Cronbach \(\alpha\) with the criterion of \(\alpha\geq 0.7\) [66]. All constructs surpass the recommended value, and therefore internal consistency can be assumed. The convergent validity was assessed following Hair et al [20] using the following 3 criteria: (1) the significance of the factor loadings, which exceed the criterion value of 0.5; (2) the average variance extracted (AVE) should be greater than 0.5; (3) the composite reliability (CR) should surpass the minimum threshold of 0.6. All subscales met these 3 criteria.

Discriminant validity was evaluated by the Fornell-Larcker Criterion [20,67]. For each latent variable, the square root of AVE (diagonal elements) must be larger than the correlation between this latent variable and any other latent variable (off-diagonal elements). As shown in Table 3, this criterion was fulfilled for all latent variables.

To further assess the quality of the structural model, we computed overall measures of goodness of fit, following the recommendations of Hair et al [20], and calculated the model chi-square statistics, the comparative fit index (CFI), and the root-mean-square error of approximation (RMSEA). Specific thresholds for high model complexity (\(\geq 30\) observed variables) and small sample size (\(\leq 250\) observations) apply. The fit indices, their values, and the specific threshold values are presented in Table 4.
Table 2. Quality criteria of the constructs.

<table>
<thead>
<tr>
<th>Latent variable and item</th>
<th>Mean (SD)</th>
<th>Standardized factor loading</th>
<th>AVE&lt;sup&gt;a&lt;/sup&gt;</th>
<th>CR&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Cronbach α</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AP&lt;sup&gt;c&lt;/sup&gt;</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AP01</td>
<td>3.34 (1.76)</td>
<td>0.943</td>
<td>0.918</td>
<td>0.957</td>
<td>.961</td>
</tr>
<tr>
<td>AP02</td>
<td>3.36 (1.77)</td>
<td>0.972</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CON&lt;sup&gt;d&lt;/sup&gt;</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CON01</td>
<td>4.40 (1.67)</td>
<td>0.885</td>
<td>0.951</td>
<td>0.951</td>
<td>.951</td>
</tr>
<tr>
<td>CON02</td>
<td>5.09 (1.60)</td>
<td>0.873</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CON03</td>
<td>4.94 (1.64)</td>
<td>0.889</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CON04</td>
<td>4.64 (1.67)</td>
<td>0.923</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CON06</td>
<td>4.51 (1.68)</td>
<td>0.886</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>IU&lt;sup&gt;e&lt;/sup&gt;</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IU01</td>
<td>4.74 (1.86)</td>
<td>0.904</td>
<td>0.806</td>
<td>0.926</td>
<td>.935</td>
</tr>
<tr>
<td>IU02</td>
<td>4.68 (1.94)</td>
<td>0.889</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IU04</td>
<td>4.64 (1.90)</td>
<td>0.902</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PB&lt;sup&gt;f&lt;/sup&gt;</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PB01</td>
<td>4.25 (1.65)</td>
<td>0.838</td>
<td>0.757</td>
<td>0.949</td>
<td>.949</td>
</tr>
<tr>
<td>PB03</td>
<td>4.11 (1.62)</td>
<td>0.901</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PB04</td>
<td>4.31 (1.70)</td>
<td>0.883</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PB05</td>
<td>4.08 (1.73)</td>
<td>0.864</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PB06</td>
<td>4.01 (1.67)</td>
<td>0.903</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PB07</td>
<td>3.70 (1.64)</td>
<td>0.827</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PC&lt;sup&gt;g&lt;/sup&gt;</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC02</td>
<td>2.85 (1.59)</td>
<td>0.877</td>
<td>0.752</td>
<td>0.938</td>
<td>.938</td>
</tr>
<tr>
<td>PC07</td>
<td>2.71 (1.46)</td>
<td>0.860</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC08</td>
<td>2.76 (1.48)</td>
<td>0.873</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC09</td>
<td>3.35 (1.56)</td>
<td>0.832</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC10</td>
<td>3.07 (1.59)</td>
<td>0.891</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SN&lt;sup&gt;h&lt;/sup&gt;</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SN01</td>
<td>4.54 (1.72)</td>
<td>0.868</td>
<td>0.782</td>
<td>0.946</td>
<td>.946</td>
</tr>
<tr>
<td>SN02</td>
<td>4.52 (1.64)</td>
<td>0.853</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SN03</td>
<td>4.96 (1.81)</td>
<td>0.875</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SN04</td>
<td>4.50 (1.78)</td>
<td>0.890</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SN05</td>
<td>4.63 (1.85)</td>
<td>0.925</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TP&lt;sup&gt;i&lt;/sup&gt;</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TP01</td>
<td>4.13 (1.60)</td>
<td>0.907</td>
<td>0.819</td>
<td>0.948</td>
<td>.947</td>
</tr>
<tr>
<td>TP02</td>
<td>4.29 (1.64)</td>
<td>0.889</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TP03</td>
<td>4.20 (1.74)</td>
<td>0.902</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TP07</td>
<td>4.30 (1.75)</td>
<td>0.921</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>AVE: average variance extracted.

<sup>b</sup>CR: composite reliability.

<sup>c</sup>AP: attitude to privacy.

<sup>d</sup>CON: perceived control over personal data.
IU: intention to use.
PB: perceived benefits.
PC: privacy concerns.
SN: social norm.
TP: trust in the provider.

Table 3. Fornell-Larcker Criterion: square root of AVE\(^a\) and correlation between latent variables (off-diagonal elements).\(^b\)

<table>
<thead>
<tr>
<th></th>
<th>AP(^c)</th>
<th>CON(^d)</th>
<th>IU(^e)</th>
<th>PB(^f)</th>
<th>PC(^g)</th>
<th>SN(^h)</th>
<th>TP(^i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AP</td>
<td>0.958</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CON</td>
<td>-0.767</td>
<td>0.891</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IU</td>
<td>-0.781</td>
<td>0.770</td>
<td>0.898</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PB</td>
<td>-0.729</td>
<td>0.560</td>
<td>0.747</td>
<td>0.870</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC</td>
<td>0.640</td>
<td>-0.803</td>
<td>-0.660</td>
<td>-0.467</td>
<td>0.867</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SN</td>
<td>-0.668</td>
<td>0.753</td>
<td>0.819</td>
<td>0.487</td>
<td>-0.610</td>
<td>0.883</td>
<td></td>
</tr>
<tr>
<td>TP</td>
<td>-0.877</td>
<td>0.851</td>
<td>0.811</td>
<td>0.639</td>
<td>-0.696</td>
<td>0.690</td>
<td>0.905</td>
</tr>
</tbody>
</table>

\(^a\)AVE: average variance extracted.
\(^b\)Diagonal elements are in italics.
\(^c\)AP: attitude to privacy.
\(^d\)CON: perceived control over personal data.
\(^e\)IU: intention to use.
\(^f\)PB: perceived benefits.
\(^g\)PC: privacy concerns.
\(^h\)SN: social norm.
\(^i\)TP: trust in the provider.
\(^j\)Not applicable.

Table 4. Goodness-of-fit measures of the CB-SEM\(^a\), following the recommendations for complex models and small samples [20].

<table>
<thead>
<tr>
<th>Fit indices</th>
<th>Sample</th>
<th>Recommended cutoff criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-square (\chi^2)</td>
<td>933.148</td>
<td>(\leq) b</td>
</tr>
<tr>
<td>Degrees of freedom (df)</td>
<td>391</td>
<td></td>
</tr>
<tr>
<td>Normed chi-square (\chi^2/df)</td>
<td>2.387</td>
<td>&lt; 3</td>
</tr>
<tr>
<td>CFI(^c)</td>
<td>0.940</td>
<td>&gt; 0.93</td>
</tr>
<tr>
<td>RMSEA(^d)</td>
<td>0.074</td>
<td>Values &lt; 0.08 with CFI &gt; 0.93</td>
</tr>
</tbody>
</table>

\(^a\)CB-SEM: covariance-based structural equation modeling.
\(^b\)Not applicable; they do not have cutoff criteria. Nonetheless, they are part of the fit indices report as standard information, which is needed for the normed chi-square (which has a cutoff).
\(^c\)CFI: comparative fit index.
\(^d\)RMSEA: root-mean-square error of approximation.

All fit indices indicate a good fit. The test of overall model fit resulted in a chi-square value \(\chi^2\) of 933.148 with 391 degrees of freedom \(df\) and a \(P\) value of <.001. Because of the dependence of the chi-square statistic on sample size and model complexity, the significant \(P\) value is negligible, and the use of the normed chi-square \(\chi^2/df\) is advisable [20]. For our model, this ratio indicates a good fit with \(\chi^2/df=2.387\), which is below the threshold of 3. Furthermore, an absolute RMSEA and an incremental fit index (CFI) were calculated. Both the RMSEA (0.074) and the CFI (0.94) meet the necessary criteria for a good model fit.

Results of the Structural Model

After the fit of CB-SEM has been evaluated, we now describe the structural model in more detail. Figure 2 represents the path coefficients and the corresponding \(P\) values. We include age, gender, education, mHealth experience, and the usage of wearables as control variables to control for the variance explained by these variables.

Table 5 summarizes the detailed analysis of the path coefficients. The \(R^2\) value for the intention to use and the other \(R^2\) values (for perceived benefits, privacy concerns, and trust in the

https://humanfactors.jmir.org/2023/1/e45503
provider) exceed the cutoff value of 0.4 \cite{68} and suggest a good model fit. Consistent with our expectations, perceived benefits has a significant effect on the intention to use ($P<.001$), as well as trust in the provider ($P<.001$) and social norms ($P<.001$), supporting H1, H3, and H4. Privacy concerns do not have a significant effect on the intention to use ($P=.14$). Consequently, H2 is rejected. Perceived control over personal data has a significant effect on privacy concerns ($P<.001$) and trust in the provider ($P<.001$), while there is no significant effect on intention to use ($P=.40$). Thus, H5a is rejected, while H5b and H5c are supported. The attitude to privacy has significant effects on perceived benefits ($P<.001$) and trust in the provider ($P<.001$), thus supporting H6b and H6d. The attitude to privacy, however, has no significant effect on the intention to use ($P=.20$) as well as on privacy concerns ($P=.41$), rejecting H6a and H6c. Our model explains $R^2=79.3\%$ of the variance in our main dependent variable, that is, intention to use mHealth technologies, controlling for demographic variables and the reported usage of wearables and mHealth apps. The control variables gender ($P=.75$), education ($P=.92$), and the reported usage of wearables ($P=.24$) were not related to the intention to use, whereas age was related negatively ($P=.002$) and the experience with mHealth apps was related positively to intention to use ($P=.03$).

Figure 2. Factor relationships in the structural model. Solid lines represent statistically significant links and dashed lines represent statistically nonsignificant links. *$P<.05$. **$P<.01$. ***$P<.001$. ns: not significant.
Table 5. Path coefficients and hypothesis testing.

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Construct A → B</th>
<th>Path coefficient</th>
<th>P value</th>
<th>Supported</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1</td>
<td>PB$^a$ → IU$^b$</td>
<td>0.380</td>
<td>&lt;.001</td>
<td>Yes</td>
</tr>
<tr>
<td>H2</td>
<td>PC$^c$ → IU</td>
<td>-0.086</td>
<td>.14</td>
<td>No</td>
</tr>
<tr>
<td>H3</td>
<td>TP$^d$ → IU</td>
<td>0.342</td>
<td>&lt;.001</td>
<td>Yes</td>
</tr>
<tr>
<td>H4</td>
<td>SN$^e$ → IU</td>
<td>0.478</td>
<td>&lt;.001</td>
<td>Yes</td>
</tr>
<tr>
<td>H5a</td>
<td>CON$^f$ → IU</td>
<td>-0.078</td>
<td>.40</td>
<td>No</td>
</tr>
<tr>
<td>H5b</td>
<td>CON → PC</td>
<td>-0.758</td>
<td>&lt;.001</td>
<td>Yes</td>
</tr>
<tr>
<td>H5c</td>
<td>CON → TP</td>
<td>0.432</td>
<td>&lt;.001</td>
<td>Yes</td>
</tr>
<tr>
<td>H6a</td>
<td>AP$^g$ → IU</td>
<td>0.110</td>
<td>.20</td>
<td>No</td>
</tr>
<tr>
<td>H6b</td>
<td>AP → PB</td>
<td>-0.729</td>
<td>&lt;.001</td>
<td>Yes</td>
</tr>
<tr>
<td>H6c</td>
<td>AP → PC</td>
<td>0.059</td>
<td>.41</td>
<td>No</td>
</tr>
<tr>
<td>H6d</td>
<td>AP → TP</td>
<td>-0.545</td>
<td>&lt;.001</td>
<td>Yes</td>
</tr>
<tr>
<td>Controls</td>
<td>Age → IU</td>
<td>-0.173</td>
<td>.002</td>
<td>N/A$^h$</td>
</tr>
<tr>
<td>Controls</td>
<td>Gender → IU</td>
<td>-0.02</td>
<td>.75</td>
<td>N/A</td>
</tr>
<tr>
<td>Controls</td>
<td>Education → IU</td>
<td>0.006</td>
<td>.92</td>
<td>N/A</td>
</tr>
<tr>
<td>Controls</td>
<td>Experience with mHealth → IU</td>
<td>0.174</td>
<td>.03</td>
<td>N/A</td>
</tr>
<tr>
<td>Controls</td>
<td>Wearable usage → IU</td>
<td>0.082</td>
<td>.24</td>
<td>N/A</td>
</tr>
</tbody>
</table>

$^a$PB: perceived benefits.
$^b$IU: intention to use.
$^c$PC: privacy concerns.
$^d$TP: trust in the provider.
$^e$SN: social norm.
$^f$CON: perceived control over personal data.
$^g$AP: attitude to privacy.
$^h$N/A: not applicable. Controls are not part of the hypothesis section; consequently, there is nothing that could be supported or rejected. Nonetheless, they are part of the results.

Discussion

Principal Findings

This study examined whether the intention to use mHealth apps could be described by an extended privacy calculus model that considers social norms, perceived data autonomy, and the attitude to privacy of the user. Furthermore, we examined the influence of control variables on intention to use, of which mHealth experience and age had a significant effect. Users who already had experience with mHealth apps and were familiar with similar apps had a greater intention to use them. This has already been demonstrated in other studies [69,70]. Age had a significant inhibiting effect on intention to use, which is in line with other studies on mHealth technology [69,70].

With overall complexity similar to existing models, the suggested model explains the variance ($R^2$) in users’ intention to use mHealth apps more effectively than other reported models (where values do not exceed 0.5 [11,18] or are not reported [19]).

An important, albeit expected, finding is that the more benefits users perceive, the higher their intention to use mHealth apps. That is, if the product is perceived to be useful or if there are benefits (eg, economic or utilitarian) users value, they are more likely to use it. Unexpectedly, in the context of health insurance apps, perceptions of benefits outweigh perceived risks, which had no part in our privacy calculus. Our model suggests that this can be attributed in part to the level of perceived control over personal data or a lack thereof, which acts as a mitigating factor that reduces or increases users’ perception of risk in the context of data protection (negative path coefficient=–0.758).

That is, the more users think they are in control of their data, the less concerned they are about disclosing personal data and vice versa.

The results of this study also underline the salient role of users’ attitudes to privacy. According to the model, the more trust is placed in the provider, the more likely the mHealth app will be used. This relationship is in part explained by the trait-factor attitude toward privacy. When privacy issues are particularly important to users, trust in the provider tends to be lower (negative path coefficient=–0.545). In addition, users’ attitude
to privacy has an indirect influence on the intention to use of mHealth apps and wearables. Users’ perceptions of benefits are negatively correlated with the attention they pay to data privacy (negative path coefficient=–0.729). Thus, the more users are concerned about data privacy, the more they devalue the benefits of data-collecting technologies. This means that in the mHealth domain, benefits (eg, financial gains as in the vignette) tend to be a less compelling argument to use this technology for those who are concerned about data privacy. However, if this relationship holds for less tangible health benefits, such as more efficient treatment, better communication with medical institutions, or early detection of diseases, remains to be seen in future studies.

Finally, social norms, that is, the opinions, experiences, and recommendations of close relatives, are also influencing the intention to use mHealth apps. In fact, social norms were the strongest drivers for the intention to use mHealth technology (path coefficient=0.478) in our study. This conforms with findings from social psychological research suggesting that people tend to adopt the opinion of their peers or relatives [71]. Thus, if the social environment supports mHealth technology use, these technologies are more likely to be used.

Implications

Based on the results, there are several possibilities for health care providers to increase the intention to use mHealth apps. First, users’ perceived data autonomy could be increased by offering an easy-to-use digital infrastructure for managing personal health data, which may ultimately increase users’ intention to use the mHealth technology. Second, because users, who are concerned about data privacy, may not want to use mHealth apps (even if they benefit them), providers may want to consider new and user-friendly ways to inform about data storage and processing policies to increase trust in critical users. This could be implemented, for example, through a user-centered app design, an easy-to-comprehend text design, and a focus on transparency [40]. Finally, to increase uptake, social norms may be activated, for instance, via testimonials of satisfied users and a reward program for recommending the app to friends and family. Additionally, customer journeys may be tracked to understand and support the social dynamics underlying the use of mHealth apps during the postpurchase phase (eg, by tracking customers’ reviews, recommendations, and posts on social media) to improve the product and ultimately increase the intention to use it [72,73].

Limitations and Future Directions

This study has several limitations that must be addressed in future research. The model was tested on a German population. However, it is evident that the use of data-collecting technology and its acceptance are strongly influenced by culture [74]. Compared with other European countries, Germans are particularly careful when it comes to using personal information online [75]. Furthermore, the sample is homogeneous in that every person residing in Germany is required to have health insurance. Thus, the probability of using a health insurance app is significantly higher than for other mHealth apps. This may also be a reason for the high explained variance ($R^2$) of the model. Future studies should check the validity and generalizability across different cultural backgrounds.

There is also the limitation that the sample was relatively tech-savvy, as evidenced by the proportion of participants who reported using wearables (73/250, 29.2%), which is higher than in previous studies. For instance, in 2021, only 21% of a representative German sample reported to use wearables regularly in a survey study [76], which could raise doubts about the representativeness of the presented data. By contrast, the number of wearable users may have also increased during the COVID-19 pandemic, which generally boosted digitalization in health care [77]. Nonetheless, future studies should validate our results in representative samples.

Another limitation is that the study’s scenario involves an app from a widely known German health insurance company, which generally has a very high reputation in the German health care system and whose motivation for publishing an app may be less driven by economic concerns than that of companies in the private sector. It is thus likely that participants perceived health insurance more positively than a commercial provider of mHealth apps. Follow-up studies must show whether the model we presented also explains the usage intention of commercial mHealth apps. Further, denial of coverage is a rather unlikely scenario in the German health care system. A more realistic scenario should be used in a future study.

Hence, future research should investigate which features trigger perceived data autonomy in users to shed more light on why apps are perceived as more or less trustworthy. A mixed methods approach (eg, an interview study to generate hypotheses and a subsequent survey study to validate them) would be a first step in examining the factors influencing the effects of perceived data autonomy on the intention to (not) use mHealth apps [78]. Finally, in this study, injunctive social norms were operationalized with respect to recommendations and approval of mHealth apps by friends and families. To what extent health professionals activate injunctive social norms to increase or decrease intention to use [24] remains to be seen in future studies.

Conclusions

We showed that our model can explain the intention to use mHealth apps more effectively than previous privacy calculus models in the mHealth domain. Specifically, we were able to show that in addition to the factors related to costs and benefits included in the original privacy calculus model, the intention to use mHealth apps is influenced by 3 additional factors: (1) The perceived data autonomy has an indirect influence on the intention to use mHealth apps by reducing privacy concerns and increasing trust in the provider. (2) The trait-factor attitude to privacy explains users’ trust in the provider and shows that users who are concerned about data privacy can hardly be convinced to use mHealth apps based on their potential benefits. (3) Social norms, that is, the opinions, experiences, and recommendations shared by one’s relatives and friends, influence users’ intention to (not) use mHealth apps. Together, these findings allow health care providers to improve their products and to increase usage by targeting specific user groups.
Acknowledgments

We acknowledge support from the German Research Foundation and the Open Access Publication Fund of TU Berlin. We also thank the eV Studienwerk Villigst and the German Federal Ministry of Education and Research, who provided the doctoral scholarship (NvK) without which this research would not have been possible. We thank the health insurance company BARMER for their cooperation in conducting the survey, especially Tanja Rehr-Meyborg and Magdalene Grahl, as well as all those who participated in the study.

Conflicts of Interest

The survey was funded by BARMER. We, the authors, state that we are not in an employment relationship with BARMER nor have we accepted any other payments. BARMER had no influence on the design of the study, the questionnaire, the analysis, and the interpretation of results. The study design for execution was given by us directly to the survey agency, just as we got the data set directly from them without any interference from the health insurance company.

Multimedia Appendix 1
Case vignette.
[DOCX File, 13 KB - humanfactors_v10i1e45503_app1.docx]

Multimedia Appendix 2
Questionnaire.
[DOCX File, 16 KB - humanfactors_v10i1e45503_app2.docx]

References


Abbreviations

AP: attitude to privacy
AVE: average variance extracted
CB-SEM: covariance-based structural equation modeling
CFI: comparative fit index
CON: perceived control over personal data
CR: composite reliability
GDPR: General Data Protection Regulation
HIPAA: Health Insurance Portability and Accountability Act
IU: intention to use
mHealth: mobile health
PB: perceived benefits
PC: privacy concerns
RMSEA: root-mean-square error of approximation
SN: social norm
SNS: social network site
TP: trust in the provider
Extending the Privacy Calculus to the mHealth Domain: Survey Study on the Intention to Use mHealth Apps in Germany

von Kalckreuth N, Feufel MA

Please cite as:
von Kalckreuth N, Feufel MA
Extending the Privacy Calculus to the mHealth Domain: Survey Study on the Intention to Use mHealth Apps in Germany
JMIR Hum Factors 2023;10:e45503
URL: https://humanfactors.jmir.org/2023/1/e45503
doi:10.2196/45503
PMID:37585259

©Niklas von Kalckreuth, Markus A Feufel. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 16.08.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Usability and Credibility of a COVID-19 Vaccine Chatbot for Young Adults and Health Workers in the United States: Formative Mixed Methods Study

Rose Weeks¹,², MPH; Pooja Sangha¹, MHS; Lyra Cooper¹, MHS; João Sedoc³,⁴, PhD; Sydney White¹,², MPH; Shai Gretz⁵, MSc; Assaf Toledo⁵, PhD; Dan Lahav⁵, MSc; Anna-Maria Hartner¹, MS; Nina M Martin⁶, PhD; Jae Hyoung Lee¹, MD; Noam Slonim⁵, PhD; Naor Bar-Zeev¹, PhD

¹International Vaccine Access Center, Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States
²Johns Hopkins Center for Indigenous Health, Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States
³Stern School of Business, New York University, New York, NY, United States
⁴Whiting School of Engineering, Johns Hopkins University, Baltimore, MD, United States
⁵IBM Research AI, Haifa, Israel
⁶Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

Corresponding Author:
Rose Weeks, MPH
International Vaccine Access Center
Department of International Health
Johns Hopkins Bloomberg School of Public Health
415 N Washington Street 5th Floor
Baltimore, MD, 21231
United States
Phone: 1 443 287 4832
Email: rweeks@jhu.edu

Abstract

Background: The COVID-19 pandemic raised novel challenges in communicating reliable, continually changing health information to a broad and sometimes skeptical public, particularly around COVID-19 vaccines, which, despite being comprehensively studied, were the subject of viral misinformation. Chatbots are a promising technology to reach and engage populations during the pandemic. To inform and communicate effectively with users, chatbots must be highly usable and credible.

Objective: We sought to understand how young adults and health workers in the United States assessed the usability and credibility of a web-based chatbot called Vira, created by the Johns Hopkins Bloomberg School of Public Health and IBM Research using natural language processing technology. Using a mixed method approach, we sought to rapidly improve Vira’s user experience to support vaccine decision-making during the peak of the COVID-19 pandemic.

Methods: We recruited racially and ethnically diverse young people and health workers, with both groups from urban areas of the United States. We used the validated Chatbot Usability Questionnaire to understand the tool’s navigation, precision, and persona. We also conducted 11 interviews with health workers and young people to understand the user experience, whether they perceived the chatbot as confidential and trustworthy, and how they would use the chatbot. We coded and categorized emerging themes to understand the determining factors for participants’ assessment of chatbot usability and credibility.

Results: In all, 58 participants completed a web-based usability questionnaire and 11 completed in-depth interviews. Most questionnaire respondents said the chatbot was “easy to navigate” (51/58, 88%) and “very easy to use” (50/58, 86%), and many (45/58, 78%) said its responses were relevant. The mean Chatbot Usability Questionnaire score was 70.2 (SD 12.1) and scores ranged from 40.6 to 95.3. Interview participants felt the chatbot achieved high usability due to its strong functionality, performance, and perceived confidentiality and that the chatbot could attain high credibility with a redesign of its cartoonish visual persona. Young people said they would use the chatbot to discuss vaccination with hesitant friends or family members, whereas health workers used or anticipated using the chatbot to support community outreach, save time, and stay up to date.
Conclusions: This formative study conducted during the pandemic’s peak provided user feedback for an iterative redesign of Vira. Using a mixed method approach provided multidimensional feedback, identifying how the chatbot worked well—being easy to use, answering questions appropriately, and using credible branding—while offering tangible steps to improve the product’s visual design. Future studies should evaluate how chatbots support personal health decision-making, particularly in the context of a public health emergency, and whether such outreach tools can reduce staff burnout. Randomized studies should also be conducted to measure how chatbots counteracting health misinformation affect user knowledge, attitudes, and behavior.

(JMIR Hum Factors 2023;10:e40533) doi:10.2196/40533

KEYWORDS
COVID-19; chatbot development; risk communication; vaccine hesitancy; conversational agent; health information; chatbot; natural language processing; usability; user feedback

Introduction

The internet’s continual availability, breadth of coverage, interactivity, and anonymity has made it a preferred health information source [1]; however, it has also propagated the spread of scientifically inaccurate, false, or misleading health information [2-4]. The COVID-19 pandemic has taken an enormous toll on human health and social functioning, raising novel and substantial challenges in communicating reliable and dynamically changing health information to a broad and sometimes skeptical public [5-9]. Although COVID-19 vaccines are thoroughly studied, misinformation abounds and is widely shared [10]. A survey in May 2021 of over 5 million Americans found adults aged 18-34 years had the highest rates of vaccine hesitancy [11], with this and other studies citing concerns regarding vaccine development, safety, and effectiveness [12-14]. This has hampered vaccine uptake in the United States, which experienced extraordinarily high COVID-19 mortality relative to other high-income countries [15].

We sought to provide access to reliable, relevant, and up-to-date information through the development of an automated dialog system, or chatbot, which supported direct questioning and engagement by users on their own terms and in their own words. Chatbots were seen early in the pandemic as a promising technology to reach and engage populations [16,17]. Chatbot performance has improved enormously in recent years, and they provide individuals with support on diverse health issues from depression to weight management [18,19]. Mental health chatbots have been shown to improve self-reported measures of depression [20]. Very limited evidence points to the potential health impact of providing vaccine information through chatbots. Experiments with crowd workers indicate that time spent engaging with a chatbot may be related to improved outcomes such as attitudinal changes related to vaccine acceptance, a promising finding that argues for making chatbot platforms compelling and engaging to incentivize chatting long enough for the intervention exposure to be sufficiently meaningful [21,22]. Chatbots must be seen by their intended users as highly usable and credible. Usability describes the effectiveness, efficiency, and satisfaction with which targeted users complete tasks on a tool in a specific context [23,24]. Credibility reflects a combination of integrity, dependability, and competence [25,26]; users judge a website’s credibility by assessing its origins, content, context, functionality, and design [27,28].

At the cusp of COVID-19 vaccine authorization for US adults aged over 18 years, we developed a web-based chatbot with an illustration of a smiley emoji in warm orange and yellow tones, embodying a friendly bot presenting credible facts about COVID-19 vaccines [29-31]. The chatbot, called Vira, short for Vaccine Information Resource Assistant, was available on a website, accessible on WhatsApp, and embedded in several other websites, such as city health departments, via an embed code snippet. IBM Research developed and managed the chatbot’s backend, which is based on a neural model that maps each user utterance to (at most) one concern from a predefined list of concerns, referred to as Key Points. Key Points were identified through various means: using a Twitter analysis, reviewing audience questions in Zoom-based public forums hosted by authors’ affiliated academic centers, and synthesizing web pages with frequently asked questions [32-34]. In addition to surfacing emerging concerns from the logs, the backend used Key Point Analysis, a commercially available technology that facilitates extractive summarization to process numerous comments, opinions, and statements and reveal the most significant points and their relative prevalence [35,36]. Vira was initially trained to respond to 100 Key Points with up to 4 alternative responses per concern to minimize repetition and enhance the naturalness of Vira’s dialog [37].

To investigate Vira’s reception with our targeted users, we sought to understand their COVID-19–related concerns, experience using other chatbots, and preferences related to information seeking. Qualitative evidence describing the experiences of health consumers, particularly racial and ethnic minority women, with health chatbots and other digital health tools is limited [38-41]. Understanding people’s needs and preferences, as told in their own words, was critical to developing a human-centered platform deemed by intended users as effective and appropriate. This study, therefore, sought to (1) understand how users assessed the chatbot’s usability and credibility and describe their intention to use the chatbot and (2) apply this understanding to improve the user experience.

Methods

Recruitment

We recruited two participant groups in urban US communities: (1) individuals aged 18-28 years; and (2) health workers, who were individuals contracted or employed by health departments to encourage the uptake of COVID-19 vaccines. We posted ads on Craigslist and Twitter targeting young people and health professionals interested in health education and communication. We invited potential participants to complete a survey in English or Spanish, with a survey link distributed via social media and email. To ensure the study sample captured a diverse range of perspectives, we recruited two participant groups in urban US communities: (1) individuals aged 18-28 years; and (2) health workers, who were individuals contracted or employed by health departments to encourage the uptake of COVID-19 vaccines. We posted ads on Craigslist and Twitter targeting young people and health professionals interested in health education and communication. To ensure the study sample captured a diverse range of perspectives, we recruited two participant groups in urban US communities: (1) individuals aged 18-28 years; and (2) health workers, who were individuals contracted or employed by health departments to encourage the uptake of COVID-19 vaccines. We posted ads on Craigslist and Twitter targeting young people and health professionals interested in health education and communication.
workers in Baltimore, Charlotte, New York City, Philadelphia, and Washington, D.C., and we used snowball sampling through professional contacts to identify health workers. We sought to achieve variability along lines of race and ethnicity to represent our targeted users. For both groups, we excluded people who stated they would “definitely NOT choose to get a COVID-19 vaccine by August 2021” in a scaled response, since the chatbot aimed to target users along the vaccine hesitancy continuum excepting those refusing vaccines [42-44].

Users were invited to participate in 3 possible activities—a web-based questionnaire, Zoom-based interview, or a web-based focus group discussion—described elsewhere [30]. Participants were given US $20 Amazon e-gift cards for each study activity completed.

Data Collection and Analysis

Usability Questionnaire

To understand Vira’s overall acceptability and its ability to respond appropriately, we presented the website and chat function to target users. We invited users to complete a Qualtrics-based written consent form, followed by a web-based Qualtrics-based questionnaire, with both forms in English. The questionnaire asked participants 10 open-ended and scaled questions about COVID-19 vaccine beliefs, previous chatbot experiences, the potential use of a chatbot to seek information about COVID-19 vaccines, and anticipated barriers. The Chatbot Usability Questionnaire (CUQ) was also included, which is a validated instrument that asked 16 questions about the chatbot’s persona, chat initiation, navigation, precision, responses, and error handling rated on a scale of 1 (strongly disagree) to 5 (strongly agree) [45,46]. We summarized the survey and CUQ responses using descriptive statistics. We also analyzed themes from the open-ended questions through cross-case comparisons, grouping responses for each question and assessing similarities and differences across responses.

Although no direct comparison can be made to other chatbot assessments, we sought to make our usability assessment results understandable to those familiar with the System Usability Scale. Therefore, we calculated participant responses to the CUQ out of 64 using the formula in equation 1, then normalized to 100 [45]. Descriptive statistics of CUQ scores are presented in the results.

The CUQ calculation is as follows:

\[ X_n \text{ is the score given by the participant on the } n\text{th question and } m=6. \]

In-depth Interviews

To solicit qualitative feedback on the chatbot’s usability, credibility, and users’ intention to engage with the tool, we conducted in-depth interviews (IDIs) with health workers and young people. Interviews were conducted in 2021 from June to October via Zoom videoconferencing software (licensed account; Zoom Video Communications Inc). After obtaining verbal consent from participants, we conducted 60-minute, audio-recorded interviews, exploring if users had difficulty using the chatbot and if they could identify intended audiences and use cases for the tool. Interviewers also asked whether the chatbot seemed trustworthy and confidential and probed to understand how users reacted to the bot persona, meaning the personality the bot assumes while interacting with a user [47]. See Figure 1A for a screenshot of the website the participants reviewed.

Recorded interviews were transcribed using Temi transcription software (Temi) and uploaded to Dedoose (version 9.0.46; SocioCultural Research Consultants, LLC), a web-based qualitative data management software. A thematic codebook was developed using a deductive grounded theory approach. First, one team member created a codebook derived from the semistructured IDI guide, with team members involved in facilitating IDIs collectively updating the codebook. Then, 2 members piloted the codebook with a handful of transcripts, noting missing codes as well as coding discrepancies. Once the codebook was finalized, 2 members coded each transcript, and a third reviewed coded text segments for discordant coding. Throughout this process, memos were used to organize and document the analytic process.

During reassembly, we grouped textual excerpts related to codes, scanning segments to discover conceptual concurrence and discord between participants and participant types (eg, opinions shared by young people but not health workers). We also identified textual data reinforcing quantitative and qualitative findings from the usability questionnaire, specifically around themes of usability and credibility.
Figure 1. (A) Interface shown to participants. Study participants reviewed this user interface. (B) User interface, following study. The new interface at VaxChat.org incorporates feedback from study participants.

Ethics Approval
This formative study was approved by the Johns Hopkins Bloomberg School of Public Health Institutional Review Board (protocol number 15714).

Results

Participant Characteristics

Usability Questionnaire Characteristics
As shown in Table 1, 58 participants completed the usability questionnaire, among whom 40 (69%) were female, with 42 (72%) holding a bachelor's degree or higher, 10 (17%) having some college or an associate degree, and the remaining 6 (10%) having a high school diploma. In all, 3 (5%) participants reported being unvaccinated against COVID-19.
Table 1. Participants’ self-reported demographic characteristics.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>IDI&lt;sup&gt;a&lt;/sup&gt; (n=11), n (%)</th>
<th>Usability questionnaire (n=58), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (IDI: n=11; usability questionnaire: n=58)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>9 (82)</td>
<td>40 (69)</td>
</tr>
<tr>
<td>Male</td>
<td>1 (9)</td>
<td>16 (28)</td>
</tr>
<tr>
<td>Nonbinary</td>
<td>1 (9)</td>
<td>2 (3)</td>
</tr>
<tr>
<td>Education (IDI: n=11; usability questionnaire: n=58)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bachelor’s degree or higher</td>
<td>11 (100)</td>
<td>42 (72)</td>
</tr>
<tr>
<td>Associate degree</td>
<td>0 (0)</td>
<td>6 (10)</td>
</tr>
<tr>
<td>Some college, no degree</td>
<td>0 (0)</td>
<td>4 (7)</td>
</tr>
<tr>
<td>High school graduate</td>
<td>0 (0)</td>
<td>6 (10)</td>
</tr>
<tr>
<td>Race or ethnicity (IDI: n=11; usability questionnaire: n=57)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Indian and Alaska Native</td>
<td>1 (9)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Asian</td>
<td>1 (9)</td>
<td>8 (14)</td>
</tr>
<tr>
<td>Black or African American</td>
<td>2 (18)</td>
<td>16 (28)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0 (0)</td>
<td>5 (9)</td>
</tr>
<tr>
<td>White</td>
<td>7 (64)</td>
<td>27 (47)</td>
</tr>
<tr>
<td>Age (years; IDI: n=11; usability questionnaire: n=58)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8-24</td>
<td>3 (27)</td>
<td>30 (52)</td>
</tr>
<tr>
<td>25-49</td>
<td>7 (64)</td>
<td>28 (48)</td>
</tr>
<tr>
<td>50-69</td>
<td>1 (9)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Income (US $; IDI: n=10; usability questionnaire: n=51)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤40,000</td>
<td>5 (50)</td>
<td>24 (47)</td>
</tr>
<tr>
<td>40,001-60,000</td>
<td>3 (30)</td>
<td>10 (20)</td>
</tr>
<tr>
<td>60,001-80,000</td>
<td>1 (10)</td>
<td>7 (14)</td>
</tr>
<tr>
<td>80,001-100,000</td>
<td>1 (10)</td>
<td>9 (17)</td>
</tr>
<tr>
<td>&gt;100,000</td>
<td>0 (0)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>COVID-19 vaccination status (IDI: n=11; usability questionnaire: n=58)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vaccinated</td>
<td>11 (100)</td>
<td>55 (95)</td>
</tr>
<tr>
<td>Unvaccinated</td>
<td>0 (0)</td>
<td>3 (5)</td>
</tr>
</tbody>
</table>

<sup>a</sup>IDI: in-depth interview.

**Interview Participant Characteristics**

Out of 11 total participants, 9 (81%) were aged 18-28 years, including 4 (36%) who worked as health workers; 2 (18%) IDI participants were health workers aged >28 years; all but 2 participants identified as female (9/11, 81%); and all were previously vaccinated against COVID-19. Of these participants, 6 (55%) also completed the web-based questionnaire.

**Quantitative Results: CUQ**

We assessed the functionality or ease of navigation with the CUQ. Questionnaire results, displayed in Table 2, indicate that most participants agreed with the statement that the chatbot was “easy to navigate” (51/58, 88%) and “easy to use” (50/58, 86%), with a corresponding proportion disagreeing that it was “very complex” (47/58, 81%). Half (29/58, 50%) of the questionnaire respondents agreed that the chatbot “understood me well,” and 74% (42/57) disagreed that it would be “easy to get confused when using the chatbot.” Additionally, 91% (53/58) of respondents disagreed that “the chatbot seemed unfriendly” but only half (32/58, 55%) felt the personality was realistic and engaging. Finally, 43% (25/58) disagreed with the statement that “the chatbot seemed too robotic.”

CUQ scores, normalized out of 100, were calculated for 56 of the 58 participants; 2 participants did not complete all 16 questions within the questionnaire. The mean CUQ score was 70.2 (SD 12.1) with a median score of 70.3 (range 40.6-95.3).
Table 2. Chatbot Usability Questionnaire Results.

<table>
<thead>
<tr>
<th>Scale items</th>
<th>Respondents, n/N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Positive scale items (Strongly Agree OR Agree)</strong></td>
<td></td>
</tr>
<tr>
<td>The chatbot was easy to navigate</td>
<td>51/58 (88)</td>
</tr>
<tr>
<td>The chatbot was easy to use</td>
<td>50/58 (86)</td>
</tr>
<tr>
<td>The chatbot was welcoming during initial setup</td>
<td>45/58 (78)</td>
</tr>
<tr>
<td>Chatbot responses were useful, appropriate, and informative</td>
<td>40/58 (70)</td>
</tr>
<tr>
<td>That chatbot explained its scope and purpose well</td>
<td>37/58 (64)</td>
</tr>
<tr>
<td>The chatbot’s personality was realistic and engaging</td>
<td>32/58 (55)</td>
</tr>
<tr>
<td>The chatbot understood me well</td>
<td>29/58 (50)</td>
</tr>
<tr>
<td>The chatbot coped well with any error or mistakes</td>
<td>28/57 (49)</td>
</tr>
<tr>
<td><strong>Negative scale items (Strongly Disagree OR Disagree)</strong></td>
<td></td>
</tr>
<tr>
<td>The chatbot seemed unfriendly</td>
<td>53/58 (91)</td>
</tr>
<tr>
<td>The chatbot was very complex</td>
<td>47/58 (81)</td>
</tr>
<tr>
<td>The chatbot gave no indication as to its purpose</td>
<td>46/58 (79)</td>
</tr>
<tr>
<td>Chatbot responses were irrelevant</td>
<td>45/58 (78)</td>
</tr>
<tr>
<td>It would be easy to get confused when using the chatbot</td>
<td>42/57 (74)</td>
</tr>
<tr>
<td>The chatbot was unable to handle any errors</td>
<td>39/58 (67)</td>
</tr>
<tr>
<td>The chatbot failed to recognize a lot of my inputs</td>
<td>36/58 (62)</td>
</tr>
<tr>
<td>The chatbot seemed too robotic</td>
<td>25/58 (43)</td>
</tr>
</tbody>
</table>

**Qualitative Results: Chatbot Usability**

Interview participants described four contributing factors necessary to achieve usability, as shown in Figure 2: (1) functionality, or ease of use/navigation; (2) performance, or the chatbot’s ability to understand and accurately respond to queries; (3) response efficiency and quality; and (4) confidentiality and privacy of the tool. Participants described two primary contributing factors to achieving credibility: (1) institutional credibility and (2) chatbot persona.

**Figure 2.** Conceptual schema derived from interview participant responses.

**Functionality**

IDIs explored Vira’s website design and usability. Regarding functionality, most participants said the chatbot was easy to use. A young woman (IDI03) said, “it’s pretty simple–you can just click on the questions that pop up and see what some basic facts are.” A minor issue noted by health workers was not seeing how to initiate a chat, such as a clear button-style indication of where to click (see Figure 1A).

**Performance**

IDIs also assessed how precisely the chatbot responded. Both young people and health workers discussed Vira’s responsiveness, commenting that they could enter questions in their own words and receive varied, appropriate responses. One young health worker (IDI01) felt that the chatbot’s ability to understand “full questions” made it “very user friendly” and more human-like:
Response Efficiency and Quality

Young interview participants described multitasking and avoiding phone calls as key incentives for using Vira (see Table 3). Young people and one health worker compared the chatbot favorably to receiving advice from a medical professional, saying they felt the chatbot may be less biased and better informed on pandemic guidance and was available for free. A young participant (IDI06) said they felt less pressure with Vira to “watch my words” compared to when they talked to their doctor because their views on vaccines may not align with their doctor’s opinions.

Table 3. Enabling and hindering factors for a COVID-19 chatbot. Summary of in-depth interview (IDI) and open-ended questionnaire responses regarding anticipated benefits and barriers of chatbot use, nonspecific to the Vira chatbot.

<table>
<thead>
<tr>
<th>Question, theme</th>
<th>Explanatory quotes (quote from questionnaire or in-depth interview with participant gender, ID number, and health worker status, if applicable)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>What would motivate you to use a COVID-19 vaccine chatbot?</strong></td>
<td></td>
</tr>
<tr>
<td>Efficiency</td>
<td>“condensed place for information,” gets “straight to the point” with a “one-tap process.” (young woman, IDI07; young female health worker, IDI01; and young female health worker, IDI08)</td>
</tr>
<tr>
<td>Avoid human interaction</td>
<td>“people my age…don’t want to make phone calls…just stay with the internet without actually having to communicate with a real person.” (young female health worker, IDI10)</td>
</tr>
<tr>
<td>Confidentiality and lack of judgment</td>
<td>“getting information without judgment” (young woman, P54)</td>
</tr>
<tr>
<td>Shareable tool to persuade others to get vaccinated</td>
<td>“For younger…16-20 year olds who feel they can’t talk to anyone…chatboxes are a great tool.” (young woman, IDI07)</td>
</tr>
<tr>
<td><strong>Do you foresee any barriers to using a chatbot?</strong></td>
<td></td>
</tr>
<tr>
<td>The expectation of rigid algorithmic design</td>
<td>“You’re kind of stuck in that rigid form of traveling through whatever path they’ve created through those algorithms.” (young female health worker, IDI13)</td>
</tr>
<tr>
<td>Query misunderstanding</td>
<td>“Frustration in getting the chatbot to understand what I want to learn or need to understand.” (young woman, P85)</td>
</tr>
<tr>
<td>Generic or nonpersonalized responses</td>
<td>“I would want to ask personalized questions about my own health” (young woman, P61)</td>
</tr>
<tr>
<td>No human interaction</td>
<td>“I really only found [chatbots] useful when I was able to communicate with a live person.” (young woman, P75)</td>
</tr>
<tr>
<td>Poor design</td>
<td>“I could see users getting frustrated quickly if…the bot is hidden somewhere on the webpage” (young woman, P74)</td>
</tr>
<tr>
<td>Concerns about accessibility for older generations</td>
<td>“User interface design needs to be attractive with human [touch], images.” (female health worker, P105)</td>
</tr>
</tbody>
</table>

Confidentiality and Privacy

Most young and health worker participants said they would use the chatbot to ask about sensitive issues and keep personal data out of a more commercial digital space. Several participants called the chatbot a “safe space” to ask questions. A young woman described this concept:

A lot of young people have family members who are anti-vaxxers. Having a chat box [sic] where they didn’t have to talk to an adult who might want to know how old they are, where their parents are, [would be] a safe space, for lack of better words.
One health worker (IDI04) who had used the Vira chatbot in her work described how she positioned it in conversations with community members:

I gave them the site while they were with me and told them to go home and ask some of the questions that they thought maybe were like dumb or didn’t want to like tell me, and they felt comfortable with doing that...that really helped in that instance.

Most participants, largely aged <30 years, were aware but not particularly worried about data privacy in this environment. A young man (IDI05) said when it came to data privacy, the chatbot compared favorably with web search engines, which “are going to take that information and use it for third-party information and ads...[the chatbot] is more private than public.” A young female health worker (IDI09) said:

...some people may feel like Hopkins is trying to gather data on what people are asking, and it very well might be, and that can still be confidential and private...I think that it would keep my information private for the most part.

Qualitative Results: Credibility of Chatbot

Interview participants listed institutional credibility and chatbot persona as key factors contributing to their determination of the chatbot’s credibility. Their assessment encompassed judgments around the chatbot’s origin, how reliable the content appeared, how they felt about using it at this stage of the pandemic, and how website design influenced their decision to use the chatbot.

Institutional Credibility

Nearly all participants said the chatbot was trustworthy because it came from Johns Hopkins University, rather than from government or pharmaceutical sources. Participants noted that while the home page had a logo at the footer, the website was overall not clearly branded as a Johns Hopkins resource, which half of all participants noted as a missed opportunity to identify the resource as trustworthy. However, one young health worker (IDI08) said she did not think it was trustworthy because “we don’t know who’s sponsoring it...[and] where my information is going.”

Chatbot Persona

The chatbot’s “mascot,” as participants referred to the smiley emoji on the home page, elicited mixed responses. Young participants, including health workers aged <30 years, said it contrasted with the topic of COVID-19—sometimes favorably, other times poorly. The mascot was said to be “silly,” “goofy,” and “very happy,” which one young participant (IDI06) felt signaled that “it’s going to be a friendly chatbot.” A male participant (IDI05) agreed, saying “it’s not going to be like the news and media where it’s...doom and gloom.” However, even this participant felt the emoji mascot was “a little much,” and another young participant felt the mascot was “a little bit creepy.” A few young people compared the mascot to a Pokémon cartoon creature. Although 2 health workers did not express concern about the mascot’s credibility, half of all young participants said the design was inappropriately childish for a tool targeted toward users like them. Although many participants liked the “warm and inviting” colors, several described the website as pink, which one female participant said would be too “girly” for male users. Several participants noted the “bright” colors could be less “overpowering,” with 2 health workers suggesting the website should use lighter, cooler colors.

Several participants expressed uncertainty about the name “Vira.” Some were not sure how to pronounce it, and others said it conveyed an association with the coronavirus. When portrayed as an acronym—Vaccine Information Resource Assistant—VIRA was better understood and accepted.

Despite these issues and Vira’s somewhat “robotic” persona, as seen by persona-related usability scores in Table 2, three interview participants said that the chatbot would help allay their pandemic-related anxiety. Regarding a breakthrough COVID-19 case, a young woman (IDI03) said:

I would be very anxious and turn to something like this to find new information. It would help me calm down.

Qualitative Results: Intention to Use

Young people said the tool would help in discussions with vaccine-hesitant friends, family members, or members of their community. In all, 9 (16%) out of 58 questionnaire respondents described using the chatbot to encourage others to get vaccinated (see exemplar quotes in Table 3).

The 6 interview participants who served as health workers felt the chatbot could support their work. First, they must keep on top of emerging concerns in the community and look up new questions. Second, listening one on one to individuals was an important part of their role. As one middle-aged health worker (IDI02) said,

We just give them a support system. They feel someone is hearing them, their issues, their opinions. They want to record their information. They want to make sure that someone is listening...[and] giving them value.

Participants described addressing the public’s concerns about vaccines via phone, for instance, in contact tracing or at health fairs, with many queries collapsing into a batch of common questions. As one young female health worker (IDI09) said:

I’ve gotten really backlogged with the amount of people that have called. There’s a lot of very similar questions. Some of them can be answered by a chatbot and it would probably streamline that process.

Health workers noted the potential of a chatbot as a source to easily access up-to-date content. As one young health worker (IDI01) said, once the resource was approved by her department of health, “I’d be using it like every time I don’t know an answer or honestly...just to double-check my work.”

Discussion

Principal Findings

This study took a mixed methods approach to measure the perceived usability and credibility of a COVID-19 vaccine information chatbot with natural language processing capability
in a web-based chat environment. An ethnically and racially diverse sample of urban-dwelling young people and health workers assessed the chatbot as achieving high usability in that it was easy to use, performed well in understanding their inputs, and offered advantages over human interactions through efficiency, confidentiality, and reliability; noted usability deficits included the chatbot’s inability to personalize responses. In the domain of credibility, participants noted the institutional affiliation with Johns Hopkins as an asset and Vira’s inappropriately cartoonish visual persona as being an area for improvement. Young people and health workers, most of whom were already vaccinated, envisioned using the chatbot as a discussion aid to encourage others to seek out vaccines. Finally, interview participants offered clear guidance to comprehensively redesign Vira’s visual persona.

Vira’s usability scores compare favorably to those of other health chatbots evaluated through standardized measures. One study examining a health chatbot with majority White adults found a mean score of 61.6 out of 100, whereas an HPV vaccine counseling chatbot used by 24 mostly White young adults scored between 74 and 80 out of 100 [19,48]. Vira was very easy to use; provided useful, appropriate, and informative responses; and explained its purpose. Although only half of the respondents thought Vira understood well and coped with input errors, this is double the score seen in evaluations of other health chatbots, showing users’ expectations for response accuracy are high [49]. In interviews, participants appreciated that the chatbot—which can handle typos and shorthand (eg, “vax” for vaccine)—could understand full sentences, and they perceived a social-like encounter. In other chatbots, so-called social bonding increases user acceptability and confidence, influences persuasiveness, and alleviates anxiety [47,50]. Since users felt less self-conscious of how they phrased a question, they may have been freer to ask sensitive questions—or encourage others to do so from the privacy of their screens. The natural typing style gave the chatbot a human-like status, and the exchange became like a social interaction where there would be no real-world consequences for a perceived stupid or inappropriate question.

Users in our study noted that the chatbot was not personalized for them and did not customize responses regarding their baseline knowledge, attitudes, vaccination status, or individual health status (eg, underlying conditions). Many other health chatbots provide personalized content and conversations to improve user engagement, dialog quality, and self-reflection [51]. This is common across downloadable apps; a review of 78 health apps with chatbots noted that 60% of these included personalization features, with most (90%) apps personalizing content—some simply addressing users by name [18]. Vira’s design as a web-based app limited such functionality. However, participants cognizant of social discord around vaccination recognized the potential of the confidentiality and privacy offered by the anonymous web-based chatbot platform. In that the main challenge regarding personalization is privacy [37], this anonymity and the “safe space” offered by Vira may be weighed against personalization.

In terms of its credibility, Vira was rated as very friendly, with participants describing the tool as having the potential—with design iterations—to be a trustworthy source for information on COVID-19 vaccines, a politicized and emotive issue. Evidence supports the notion that chatbots presenting information about controversial topics can be convincing and trustworthy, especially with supporting links [52,53]. Many users of health chatbots report high satisfaction and positive perceptions, and the use of even moderately rated chatbots has been associated with behavior changes [19,20]. Although this study was not designed to evaluate the effectiveness of Vira in changing participant attitudes, their appraisal of the chatbot as being highly usable supports potential pathways toward behavior change to explore further.

Participants stated an interest in using Vira in their personal lives and, in the case of health workers, in professional roles. Large health organizations evidently understand the potential of chatbot technology, with the health care chatbot market expected to reach nearly $US 1 billion by 2027 [54]. Health agencies and some US states have also launched health chatbots during the COVID-19 pandemic to encourage widespread sharing of credible health information [55-58]. Further investments in high-performing, user-validated chatbots would aid health educators in communicating about rapidly changing vaccination guidance. In addition, installing chatbots on health department websites could reduce call volume and support public health workers [59].

This formative study provided investigators with user feedback to iteratively improve the user experience for Vira, a chatbot designed rapidly to support vaccine queries during the pandemic’s peak, including on the visual persona, provided rapid feedback for a website redesign (see Figure 1B). The new VIRA, spelled in all capital letters, is shown in calming blue and purple tones as a smaller, still-friendly orb supporting human users. The chatbot’s response database, or repository of potential responses, was also comprehensively edited via a separately reported message testing study [30].

However, numerous questions remain: What are the social implications of automating conversations about vaccine decisions, previously a person-to-person encounter highly reliant on trust? As mental health chatbots can reduce anxiety and loneliness, can vaccine chatbots simulate a support system validating people’s search for answers—helping them feel heard, even if by a bot? Evidence is needed, through a randomized evaluation, to explore which elements of a chatbot interaction, such as chat duration or added personalization, could lead to measurable changes in vaccine attitudes and behavior or, indeed, impact related to other stigmatized health issues such as sexual health. Whether chatbots effectively counter health misinformation and support the use of motivational interviewing, one of the only evidence-based means to soften vaccine hesitancy, is another important area for exploration [60,61].

Although this study provided rapid feedback to course-correct Vira’s visual design and inform its outreach strategy, it has several limitations. First, our participants were mostly college-educated, perhaps due to a reliance on Twitter ad recruitment [62]. In addition, participants were recruited using Johns Hopkins—branded ads and may have been more favorable toward the institution than others who did not click on the ads.

https://humanfactors.jmir.org/2023/1/e40533
IDI recruitment by Johns Hopkins of several health workers employed in Baltimore likely introduced bias, since participants may have been less likely to offer unfavorable comments out of a sense of collegiality; nonetheless, such participants did offer specific critiques. The conduct of the study at a university widely known to promote COVID-19 vaccination likely dissuaded some vaccine skeptics [63]. Since investigators needed to collect data rapidly to alter a tool already in use during a pandemic, we conducted a small number of qualitative interviews, focusing on professional users. Nearly all our participant sample said they had been vaccinated, which is not representative of young adults and health workers in the United States; however, a large proportion of the US public who got primary doses of COVID-19 vaccines have not subsequently obtained booster doses [64]. Therefore, we believe the results are relevant to support efforts to counteract vaccine hesitancy. Further, most participants describe theoretical usefulness in a one-time encounter with the chatbot. A final limitation is that this study describes Vira’s performance at launch; since then, the number of Key Points Vira can address has nearly doubled, and the performance of its adaptive algorithm has presumably improved.

Conclusions
Launched at the peak of the COVID-19 pandemic, as cases caused by the Delta variant crested in the US, Vira offered a highly functional and credible system to respond quickly and appropriately to users’ vaccine queries. We used rich data gathered through interviews to identify and remedy deficits in the chatbot persona. Young people and health workers in the study felt chatbots offered significant benefits in a pandemic context due to their reliability, responsiveness, and efficiency and that the Vira chatbot was a credible and private way to seek information on a sensitive issue. More research is needed to determine how guidance offered in an anonymous 2-way dialog, potentially designed to simulate a motivational interview, could shift perceptions of emotionally charged issues with participants in a real-world setting. Evidence is also needed to measure whether chatbots strengthen public education services and are cost-effective if made widely available.

Acknowledgments
This research and the development of the chatbot “VIRA” (VaxChat.org) was supported by the Johns Hopkins COVID-19 Training Initiative with funding from Bloomberg Philanthropies and in-kind support from IBM Research. Meta provided technical and advertising support enabling WhatsApp integration, not described in the manuscript.

Conflicts of Interest
RW reports support from an educational grant from Johnson & Johnson Foundation for chatbot implementation, funding which did not overlap with this study conduct and analysis. NBZ reports investigator-initiated research grants to his institution from Merck, Serum Institute of India, and the Asian Development Bank, unrelated to this work. NBZ also reports receipt of consulting fees on one occasion from Merck for providing methodological advice unrelated to this work.

References


29. VIRA is a chatbot from Johns Hopkins that answers COVID-19 vaccine questions. Vaxchat. URL: https://vaxchat.org/ [accessed 2022-03-29]


47. Fadhiel A, Schiavo G. Designing for health chatbots. arXiv Preprint posted online on February 24, 2019. [FREE Full text]


Abbreviations
- CUQ: Chatbot Usability Questionnaire
- IDI: in-depth interview
- Vira: Vaccine Information Resource Assistant

©Rose Weeks, Pooja Sangha, Lyra Cooper, João Sedoc, Sydney White, Shai Gretz, Assaf Toledo, Dan Lahav, Anna-Maria Hartner, Nina M Martin, Jae Hyong Lee, Noam Slonim, Naor Bar-Zeev. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 30.01.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
A Medical Assistive Robot for Telehealth Care During the COVID-19 Pandemic: Development and Usability Study in an Isolation Ward

Ruohan Wang1*, BSc; Honghao Lv1*, BSc; Zhangli Lu1, BSc; Xiaoyan Huang2, PhD; Haiteng Wu3, PhD; Junjie Xiong3, MSc; Geng Yang1,4, PhD

1State Key Laboratory of Fluid Power & Mechatronic Systems, School of Mechanical Engineering, Zhejiang University, Hangzhou, China
2College of Electrical Engineering, Zhejiang University, Hangzhou, China
3Hangzhou Shenhao Technology Co, Ltd, Hangzhou, China
4Zhejiang Key Laboratory of Intelligent Operation and Maintenance Robot, Hangzhou, China

*these authors contributed equally

Corresponding Author:
Geng Yang, PhD
State Key Laboratory of Fluid Power & Mechatronic Systems
School of Mechanical Engineering
Zhejiang University
38 Zheda Road
Hangzhou, 310027
China
Phone: 86 18757132667
Email: yanggeng@zju.edu.cn

Abstract

Background: The COVID-19 pandemic is affecting the mental and emotional well-being of patients, family members, and health care workers. Patients in the isolation ward may have psychological problems due to long-term hospitalization, the development of the epidemic, and the inability to see their families. A medical assistive robot (MAR), acting as an intermediary of communication, can be deployed to address these mental pressures.

Objective: CareDo, a MAR with telepresence and teleoperation functions, was developed in this work for remote health care. The aim of this study was to investigate its practical performance in the isolation ward during the pandemic.

Methods: Two systems were integrated into the CareDo robot. For the telepresence system, a web real-time communications solution is used for the multiuser chat system and a convolutional neural network is used for expression recognition. For the teleoperation system, an incremental motion mapping method is used for operating the robot remotely. A clinical trial of this system was conducted at First Affiliated Hospital, Zhejiang University.

Results: During the clinical trials, tasks such as video chatting, emotion detection, and medical supplies delivery were performed via the CareDo robot. Seven voice commands were set for performing system wakeup, video chatting, and system exiting. Durations from 1 to 3 seconds of common commands were set to improve voice command detection. The facial expression was recorded 152 times for a patient in 1 day for the psychological intervention. The recognition accuracy reached 95% and 92.8% for happy and neutral expressions, respectively.

Conclusions: Patients and health care workers can use this MAR in the isolation ward for telehealth care during the COVID-19 pandemic. This can be a useful approach to break the chains of virus transmission and can also be an effective way to conduct remote psychological intervention.

(JMIR Hum Factors 2023;10:e42870) doi:10.2196/42870

KEYWORDS
COVID-19; MAR; telehealth care; video chat system; mental health care
Introduction

Background
The COVID-19 pandemic has been affecting the global population for more than 2 years since the World Health Organization’s declaration of its outbreak on March 11, 2020 [1]. Despite being first and foremost a health crisis, COVID-19 has the seeds of a mental health crisis [2,3]. People feel frustrated, worried, and stressed, not only due to the immediate health impacts of the virus but also due to the lack of social communication caused by movement restrictions [4-7]. In face of the pandemic, one major solution to reduce the spread of the virus is keeping social distance [1,8], which means less physical contact and even physical isolation. The era of smart medicine, known as Healthcare 4.0, makes medical care more efficient and intelligent. Healthcare 4.0 is leading to a revolution in health care services to cope with global medical challenges, especially in isolation care, in which telehealth assistance can be deployed [9,10]. Telehealth assistance allows health care workers to implement medical treatment without contact with patients, directly breaking the transmission chains of the virus [11-13]. One of the paradigm shifts in telehealth is the communication model from direct consultation to human-computer contact, in which a medical assistive robot (MAR) can be adopted as a critical way for delivering clinical mental health care to relax nervous individuals during this crisis [14,15].

Prior to the COVID-19 pandemic, the most important application of MARs was robotic surgery [16]. During the pandemic, MARs have proliferated for contactless medical care purposes. Organizations such as the World Health Organization and the Centers for Disease Control and Prevention utilize MARs for suggestion-giving, emotion-guiding, and information-sharing applications [17,18]. Moreover, to achieve a more intimate interaction, the appearance of MARs is developing toward humanoid robots. Teleoperation and telepresence functions are also integrated into the robots, allowing them to move around and monitor the patient in the isolation ward [19]. MARs, if effectively designed and used, can bridge the gap between patients and telehealth care providers during the pandemic.

This paper describes a robot named CareDo with remote chatting, facial expression recognition, and teleoperation functions for telehealth care in the isolation ward, aiming to provide a safe and efficient interaction between patients and doctors during the pandemic. Figure 1 shows the overall structure of the proposed CareDo robot system. The robot is equipped with a moveable chassis, a collaborative robot (YuMi), a camera for facial expression recognition, a microphone for voice input, and a customized tablet for video chat. Both the patient and robot are in the isolation ward, and the patient can use voice commands to wake up the video chat system equipped on the robot. Once other users log in to the system, they can conduct multiuser real-time video conversations.

Figure 1. Overall architecture of the proposed medical assistive chatbot system for telepresence and telehealth care.

The primary contributions and novelties of this work are as follows: (1) an advanced MAR was developed and integrated with voice command interaction and human motion–based teleoperation; (2) a multiuser video chat system based on web real-time communication (WebRTC) was deployed with the facial expression recognition system using a trained convolutional neural network (CNN) model; and (3) a voice activation detection algorithm was designed and used during the voice command interaction, which is self-adaptive to the environment sound intensity and significantly improved voice recognition accuracy.

Related Works
The past 2 years have witnessed an increasing number of robots in hospitals. Robots are considered to be an effective tool for cutting off the transmission of the virus.

Various robotic solutions have been implemented for reducing unnecessary physical contacts in coronavirus management. Representative robots used for these purposes are presented in Figure 2.

A new hospital in Wuhan, China, adopted robots to deliver food, drinks, and drugs to patients in the initial stage of the COVID-19 epidemic [20]. Some of these robots are humanoid with wheeled bases and move semiautomatically in the hospital controlled...
by the medical staff. TIAGo [21,22], a robot operating system (ROS)-based robot platform, can perform both grasping tasks and disinfection tasks automatically. Users can choose the operating model for two scenarios through web graphical user interfaces (GUIs). Moxi [23], a robot with similar functions as the TIAGo robot, can perform repetitive chores such as grasping, pulling, opening, and guiding objects for hospital staff. Similar to the two robots described above, Lio-A also has a single arm and is able to move autonomously [24]. Moreover, Lio-A, equipped with loudspeakers and a multidirectional microphone, can understand some commands and interact with humans. Lio-A has a display screen on its front, which can show the text during its voice interaction with a human.

Figure 2. Representative medical robots used in hospitals during the COVID-19 pandemic. (a) Vici robot developed by InTouch Health (United States). (b) Moxi robot created by Diligent Robotics (United States). (c) Lio-A robot from F&P Robotics AG (Switzerland). (d) TIA-Go robot from PAL Robotics (Spain).

The assistive robots mentioned above are mainly used for logistics and disinfection. To provide emotional care, tools with human-machine interaction capacity are being released. For instance, Podrazhansky et al [25] developed a system for conducting surveys and retrieving health data. El Hefny et al [26] proposed a character-based virtual robot for reducing the risk of misinformation amplification. Amer et al [27] presented a chatbot system that can answer questions related to COVID-19. These human-machine interaction systems might lack human empathy [26]. Hence, a chatbot specially designed as a humanoid model has been proposed to improve the above systems [28]. Vici is a robot located in a hospital for telehealth [29]. Doctors can communicate with patients using the diagnostic function on Vici without direct patient contact. Pudu is a social robot for communication and telepresence functions, which can be remotely controlled using its teleoperation mode [30]. Medbot delivers telehealth in India by answering patients’ questions about health care, including home remedies, local food diets, and the detection of common diseases [31].

From these current related works, it can be seen that MARs are becoming ubiquitous, especially during the pandemic when people’s movement has been restricted. Nevertheless, a previous study showed that there are potential safety issues when using conversational assistants for health information purposes [32]. According to the benefits and drawbacks of the above accomplishments, this study considered the needs of patients, health care workers, and various application scenarios during the COVID-19 pandemic in designing the CareDo MAR.

Methods
System Architecture
The CareDo MAR used in this work includes two main functional parts: a telepresence system and a teleoperation system. The telepresence system contains two subsystems: a multiuser video chat system and a facial expression recognition system. The former allows the patient to talk with doctors or families without physical contact, whereas the latter can be used for patient emotional monitoring. The teleoperation system is a supported physical assistance solution for noncontact telehealth care. In the teleoperation system, the main technology is the motion mapping method, which was introduced previously [33]. With the two functional parts mentioned above, this assistive robot can be regarded as the second body of medical staff. Hence, CareDo incorporates relevant methods of a telepresence system and its novel application strategies in assisting with a teleoperation system.

As shown in the schematic diagram of the system in Figure 3, three elements, the doctor/health care worker, the patient, and the robot, are involved. The MAR, acting as a telehealth care task performer in this system, is located in the isolation ward and controlled by the health care worker in the call center of the hospital. In this way, physical contact between the patient and medical staff is blocked. The two systems equipped on the robot play an important role in the enhanced interaction between the doctor and patient. From the site of the health care workers, the patient can receive assistive behavior, traditionally
completed through psychological intervention and physical assistance, based on the teleoperation system. In addition, vital signs and the emotional status of the patient can be obtained by the doctors via the telepresence system. The dual-arm robot YuMi was chosen as the manipulator for carrying out physical assistance for the patients [34]. This system is a multinode distributed control system based on an ROS. The details are provided in the following sections.

**Figure 3.** Detailed teleoperation and telepresence systems diagram of the proposed medical assistive chatbot. API: application programming interface; WebRCT: web real-time communication.

**WebRTC-Based Video Chat System**

A self-developed video chat system is integrated on the dual-arm MAR for chatting function realization. Two main technologies, a real-time speech recognition function and a noncontact telepresence GUI, are used on the video chat system of the robot. With these technologies, the robot can act as a medium for remote consultations and video chats.

For the real-time speech recognition component, the chat system is designed to recognize the voice input of a video opening construction. Toward this end, pocketsphinx, an offline voice recognition package with a specific speech recognition acoustic model, is integrated in the robot to handle the voice input. The voice activation detection algorithm is used to enable the robot to start sound recording and the recording ends with the last word. To extract voice information from the audio information, a threshold-based decision criterion is used. When the surrounding sound is stable, it has a sound energy denoted as $E$. The threshold value $\varepsilon$ is then obtained using a previously described data preprocessing method [35]. The threshold $\varepsilon$ represents the voice energy needed to trigger the voice recording process. In an isolation ward, the level of noise typically fluctuates because of the operation of various medical instruments. Hence, the trigger threshold $\varepsilon$ was set to be self-adaptive to the environment sound intensity. Assuming $\varepsilon=f(E)$, where $\varepsilon \in \{E_{\text{min}}, E_{\text{max}}\}$, through data set preprocessing, the threshold values $\varepsilon_{\text{min}}$ and $\varepsilon_{\text{max}}$ can be set using the sound energy $E_{\text{min}}$ and $E_{\text{max}}$ respectively. The self-adaptive threshold value can then be expressed as:

In the sound recognizing and matching process, a pretrained dictionary file is used to save the related words about logging the live video chat GUI. Then, the pocketsphinx package will find the parameter with the most similar meaning and obtain the final recognition results to determine whether to open the GUI by comparing the input voice signal and characteristic parameter in the template library.

The noncontact telepresence GUI was designed to offer a multiperson remote video platform for patient condition consulting and chatting. Therefore, the WebRTC communication technology [36] was used on the robot chat system to realize the transmission of video/audio streams. WebRTC allows network sites to establish peer-to-peer connections between browsers without intermediate media. Moreover, to go beyond a simple one-to-one video call, multiple RTCPeerConnections are used on WebRTC to offer connections for every endpoint to every other endpoint in a mesh configuration.

The entire video chat system structure is schematically presented in **Figure 4**. The system uses the voice input method mentioned above to extract the human voice from environment noise and to detect whether people have finished speaking. The most frequently used voice commands designed for the current use cases in the hospital are listed in **Table 1**. For all commands, 2 to 4 keywords of each chatting stage were set up to improve the reliability of speech recognition. In practical usage, the chat system of CareDo has the ability to distinguish the patient’s
voice commands for contacting different doctors. Various approaches were utilized to achieve this function: (1) information of the related doctors was added to a contact list inside the robot system, enabling patients to send voice commands (including the doctor’s name) to contact an appointed doctor, and (2) doctors with different responsibilities were assigned unique numbers so that the patient can speak the voice commands with the doctor number and then the robot can directly contact the responsible doctor. In addition, as shown in Table 1, combined with a duration varying from 1 to 3 seconds of each common command, the voice activation detection algorithm is optimized and improved for enhancing the sensitivity of voice command detection. Once speaking is finished, the voice will use the online Baidu application programming interface for recognition. On the one side, the recognition results will trace back to the local computer, whereas on the other side, the voice constructions enter into the WebRTC Video & Audio System on which the consultation system GUI is built. Health care workers or families can take video calls with the patient through the GUI remotely to consult on the patient’s physical and mental health.

**Figure 4.** Flow-process diagram of the audio and video system for remote consultation. API: application programming interface; WebRTC: web real-time communication.

**Table 1.** Customized voice commands for the designed web real-time communication–based video chat system.

<table>
<thead>
<tr>
<th>Chat stage and command instructions</th>
<th>Approximate duration (seconds)</th>
<th>Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Wakeup system</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“Hi, CareDo”</td>
<td>1</td>
<td>Hi; Hey; Hello; CareDo</td>
</tr>
<tr>
<td>“Start remote consultation”</td>
<td>3</td>
<td>Start; Consultation</td>
</tr>
<tr>
<td><strong>Video chat</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“Create a meeting room”</td>
<td>3</td>
<td>Create; set</td>
</tr>
<tr>
<td>“Enter the meeting room”</td>
<td>2</td>
<td>Enter</td>
</tr>
<tr>
<td>“Call Doctor Wang”</td>
<td>2</td>
<td>Call; Doctor</td>
</tr>
<tr>
<td><strong>Quit system</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“Exit meeting room”</td>
<td>3</td>
<td>Exit; Quit</td>
</tr>
<tr>
<td>“Thanks, CareDo”</td>
<td>2</td>
<td>Thanks; CareDo</td>
</tr>
</tbody>
</table>

The proposed system can realize remote consultation as well as daily family chats without health care workers entering the isolation ward. Users can log in to this video chat system through general desktop browsers such as Google Chrome and Microsoft Edge. They do not need to download specific software. Therefore, this system is safe because of reduced exposure to any vulnerabilities that may exist on the vendor’s client.

**CNN-Based Facial Expression Recognition**

In addition to the remote video chat system, a facial expression recognition system based on CNN is used to monitor the emotional fluctuation of the patient, providing retraceable historical data for intervention therapy and promoting patients’ mental health as well as disease management. This system was achieved from our previous work on facial expression recognition for human-robot interactions [37]. **Figure 5 shows**
the process of facial expression recognition. The source images for recognition are provided by the camera mounted on the robot. Since the source image contains some nonfacial regions, the face detection algorithm is used for detecting the region of the human face. Because of the differences in the size, aspect ratio, and illumination conditions of images, facial image preprocessing needs to be implemented to unify these image features. Measures such as image cropping, resizing, and normalizing are used to preprocess the image to remove some irrelevant information of the face region, distinguish more subtle facial information, and adjust the image size. Furthermore, random flip technology is used for removing high-frequency noise and insuring a similar distribution of the image pixels. Following image preprocessing, the CNN-based network is used for facial expression decoupling. The generative and discriminative representations are learned simultaneously. A classifier was developed by training the features obtained in the last step using a machine learning algorithm. The data set Fer2013, which consists of 35,887 grayscale images of faces with emotion, was used for training the model, as shown in Figure 5. A detailed description of the model architecture was provided previously [38]. The first 32,299 images in Fer2013 were used as the training sets and the remaining 3587 images were selected as the verification sets. For model training, we used the configuration of the 50,000 training steps with a learning rate of 0.0001. Finally, the facial recognition result is obtained through the processes mentioned above. Five common facial expressions were defined and classified in this work: neutral, surprise, sad, fear, and happy.

**Figure 5.** Instruction for the facial expression recognition process.

---

**Human-Cyber Physical System–Based Remote Assistive Technology**

**System Structure**

To assist patients in the isolation ward, a unique teleoperation system is proposed to provide an intuitive remote-control interface for doctors to operate the MAR. As a human-cyber physical system (HCPS)-based assistive technology, three elements are included in this system. Health care workers, as the humans in this system, wear a motion capture device suit. The MAR, as the physical entity in this system, can be remotely controlled by health care workers [38]. The cyber can be the information transferred from the human side to the robot side, where physical interventions on the patient can be implied. According to the detailed control block diagram of the system shown in Figure 6, the proposed telerobotic system can be divided into a motion-capture subsystem on the operator site and a robot-control subsystem on the robot side.
Figure 6. Illustration and use case of the human-cyber physical system–based remote assistive technology.

Human Side
The human motion capture technology is mainly used on the human side of the teleoperation system. The Perception Neuro 2.0 (PN2) motion capture suit is used to capture the real-time upper limb motion of the operator. PN2 is an adaptive motion capture device that consists of multinode inertial measurement units (IMUs), which are all located on the straps in this device [39]. IMUs can transmit the heading angle, acceleration, and angular velocity information to the hub, which is the central processing unit of PN2. However, different wearers have distinct body sizes. Therefore, to obtain the position and orientation information of the hand IMU relative to the hip IMU of each wearer, the parameters of the body parts such as arm length and shoulder width must first be measured and input into the Axis Neuron software. In addition, a self-developed executable program is used to obtain the motion tracking data from Axis Neuron, a supporting application of PN2, and communicate with the ROS. In the ROS, two nodes are established to receive and publish the motion data of the limbs and hands.

Robot Side
From the human side mentioned above, the position and posture data of the operator hands are obtained. Because the workspace of a human hand and the robotic manipulator is different, a previously proposed incremental pose-mapping strategy was used [33]. This method is mainly used to obtain the current human hand orientation and the increment of its position, and then to map it to the robot based on the current position of the robot. Using the open-source inverse kinetic algorithm trac_ik [40], each joint angle of the dual arm can be obtained corresponding to the current robot pose. The predefined different hand gestures stand for different robot motion control commands. Based on these, Lv et al [33] developed a hybrid mapping method of hand gestures and limb motion. Before the teleoperation begins, the operator does not need to assume the same posture as the robot arm. Hand gestures can be defined to enable and disable motion mapping. Hence, on the human side, the action of the operator can be more flexible, while on the robot side, the manipulator can reach any position in its workspace.

Ethics Considerations
Approval of all ethical and experimental procedures and protocols was granted by the Clinical Research Ethics Committee of the First Affiliated Hospital, Zhejiang University (FAHZU; approval number IIT20200048A-R1), and the study was performed in line with the full informed consent of the volunteers, in accordance with all local laws.

Results
Performance of the WebRTC-Based Video Conference System
The WebRTC-based video chat system provides essential telemedicine services. Compared with other video chat systems, this system adopts peer-to-peer connection, which is easy to manage and deploy. Figure 7 shows a practical use case of the video chat system on both a computer and a mobile phone and a video presentation of this use case is provided in Multimedia Appendix 1.
The GUI on the browser is used as shown in the window on the left side of Figure 7. Integrated with the speech recognition function, the video chat system can be awakened and controlled by the voice command from the user, both from the patient side and the remote doctor side. This technique enables the noncontact interaction between the robot and patients, which decreases the cross-infection risks for doctors and other medical staff when they operate the robots. In this test case, three subjects were in different locations and used different local area networks to log in to the system at the same time. Two subjects entered the system by using the voice wakeup function and they then launched the video and voice applications for communication. One subject opened the remote screen through which the medical instructions or psychological counseling methods were shared. The text window chat function was tested for transferring text messages and medical documents. This GUI was tested on both a computer and a mobile phone. The tests confirmed its usability in this video chat system.

**Facial Expression Recognition Performance**

The facial expression recognition performance was evaluated during clinical trials at the FAHZU. Using the camera integrated on the front screen of the CareDo robot, the facial expressions of the patients were recorded and analyzed. The facial expression recognition data set was collected from 12 subjects, including 6 female and 6 male subjects, ranging in age from 15 to 60 years and evenly distributed from three groups: the young group (15-25 years old), middle-aged group (25-45 years old), and older adult group (45-60 years old). The facial expression recognition system worked 8 hours a day and each subject’s facial expression was recorded 15 times. The final data set was composed of 180 records (12 subjects×15 times/subject), among which 152 valid records were obtained. Table 2 shows the verification results of the facial expression recognition of one patient in a single day using the CareDo robot in the isolation ward. Cross-validation was conducted for the recorded expressions of the patients and the recognition accuracy is provided in the table for all five expression types. From the validation results, we can easily see that the neutral facial expression was detected as the most common emotion for this test, followed by the happy facial expression. The highest recognition accuracy reached 95% for the happy expression.

**Table 2.** Accuracy of recognition of patient facial expressions.

<table>
<thead>
<tr>
<th>Facial expressions</th>
<th>Records in one day (times)</th>
<th>Recognition accuracy, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surprise</td>
<td>0</td>
<td>Not applicable</td>
</tr>
<tr>
<td>Sad</td>
<td>5</td>
<td>80.0</td>
</tr>
<tr>
<td>Neutral</td>
<td>125</td>
<td>92.8</td>
</tr>
<tr>
<td>Fear</td>
<td>2</td>
<td>50.0</td>
</tr>
<tr>
<td>Happy</td>
<td>20</td>
<td>95.0</td>
</tr>
</tbody>
</table>
Verification in the FAHZU Emergency Intensive Care Unit

According to the current diagnosis and treatment operation requirements of the isolation ward for COVID-19 patients, we developed a new type of the CareDo MAR to assist in the diagnosis and treatment operations of medical staff. After validation in a laboratory environment, the robot was checked by the clinical research ethics committee of the FAHZU and obtained investigator-initiated trial (IIT) ethics approval. The CareDo robot was then applied in the emergency intensive care unit (EICU) of the FAHZU for preliminary clinical function verification, as shown in Figure 8.

Aiming at reducing the risk of infection to health care staff due to exposure to the COVID-19 virus, the robot was used in the isolation ward to perform remote care tasks through teleoperation. For the WebRTC-based video chat system, the COVID-19 patient interacted with the remote doctors using the interactive screen on the front of the robot. As shown in Figure 8a, using voice and video interactive devices, doctors can chat with the patient and perform some routine diagnoses remotely. In addition, the mental health status of the quarantined patients in the isolation ward would be a greater concern than that of general patients. Hence, with use of the facial expression recognition system, the CareDo robot acts as the bedside companion of the patient by observing the patient’s facial expression status, as shown in Figure 8b. The doctor can then communicate with the patient remotely to provide any psychological intervention guidance according to the results and analysis of facial expression recognition. The interactive screen can also play some related informational and educational videos for patients (Figure 8c). For the remote assistive system, the CareDo robot was teleoperated to perform some medical delivery tasks using the proposed HCPS-based remote assistive technology, as shown in Figure 8d-f. The robot in the teleoperation function can be used for delivering medicine or medical supplies such as a thermometer, food, personal supplies, and other required items to patients. Other details about the implementation of the MAR in the FAHZU were reported in our previous paper [13]. In summary, the developed CareDo robot has been applied in real isolation wards with the video chat system and the remote assistive system. All of the desired functions have been preliminarily achieved based on the basic requirements of both doctors and patients, and positive feedback from the users has been reported in the real clinical trials in the FAHZU.

Figure 8. Application cases of the CareDo assistive robot used in the First Affiliated Hospital, Zhejiang University emergency intensive care unit during the COVID-19 pandemic. (a) Voice and video interaction between the doctor and patient. (b) Facial expression recognition. (c) Educational videos to provide information to the patient. (d-f) Remote medical delivery tasks delivered via teleoperation.

Discussion

Improvement of Telehealth Services

The emergence of COVID-19 has brought great changes to the medical industry, especially the telemedicine service. The CareDo robot offers another new form of telehealth assistance. In this work, an advanced telerobotic system was developed. Its efficient deployment in hospitals was applied by leveraging the enabling technologies of Healthcare 4.0. Techniques, including high-performance wireless communications, high-quality remote audio and video systems, an intelligent remote-controlled robot, and wearable sensors for motion capture, are used to assist and protect health care professionals. With these functions, CareDo can execute relevant operations of a remote video system according to the patient’s voice instruction, monitor patients’ mental health status, and grasp and deliver medical supplies through teleoperation. During the utilization period in a hospital, CareDo can mitigate the risk of nosocomial infection and therefore contribute to accelerated recovery of the COVID-19 epidemic.

In the proposed telerobotic system, telemedical staff can use remote video to talk with patients and remotely operate the robot outside the negative pressure ward to complete nursing work, avoiding cross-infection caused by their frequent close contact with patients. The proposed system can realize the real-time monitoring and recording of patients’ emotional changes.
providing retraceable historical data for intervention therapy and promoting patients’ mental health as well as disease management. The system makes significant contributions to the mitigation and suppression of COVID-19 transmission chains for impacted societies.

Limitations and Future Work

This effort offers a quick solution of remote video and dialogue between patients and doctors during the pandemic. However, several limitations still exist. First, the user experience has not been deeply investigated or estimated during the use of a single function such as a video chat. Second, for the telepresence system, the recognition accuracy of facial expressions such as sad and fear still need to be optimized. A longer patient usage time is suggested to obtain more samples and records. During the implementation and clinical trials, the influence of wearing a mask was not considered or tested in this work. Wearing a mask will cover the lower part of the face and make most facial features invisible, which will decrease the facial expression recognition accuracy [41,42]. Third, for the teleoperation system, this work was based on unilateral teleoperation and we did not investigate the force feedback from the robot to the operator. Furthermore, the dual arms are controlled by human hands, lacking consideration of cooperation tasks. More complex tasks and flexible control methods should be considered to achieve compliance control.

Future work could focus on the improvement of functionality and integration. Based on the exploitable functions of WebRTC, the facial expression recognition function can be integrated into the real-time communication system. The influence of wearing masks on facial expression recognition will be considered and investigated in the future. The user operation process can be simplified while the security of the remote video chat system can be evaluated. Further study can also focus on developing cost-effective MARs for applications in more generalized telehealth care scenarios.

Conclusions

This article described the design and development of CareDo, a MAR devised to provide telehealth care to COVID-19 patients in the isolation ward. Three key technologies used on this robot are (1) a telepresence system in which the user can log in with voice input, enabling patients, doctors, and patients’ family members to have safe and real-time remote chats; (2) a facial expression recognition system that can monitor the patients’ emotional fluctuations; and (3) multinode ROS-based teleoperation technology that assists the robot in the isolation ward to perform other tasks such as delivering medical supplies. The CareDo robot was used in the EICU of the FAHZU for function verification under IIT ethics approval. The results showed that use of this MAR in the hospital can reduce the risk of cross-infection between patients and doctors. Moreover, the multiuser video chat system allows patients to talk with doctors and their family, which can relieve the patients’ mental stress from isolation.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under grant agreements 51890884 and 51975513, Natural Science Foundation of Zhejiang Province under grant agreement LR20E050003, Major Research Plan of Ningbo Innovation 2025 under grant agreement 2020Z022, and in part by the Zhejiang University Special Scientific Research Fund for COVID-19 Prevention and Control under grant 2020XGZX017.

Authors’ Contributions

GY provided directional guidance for the research. RW and HL conceived and designed the experiments. ZL performed the experiments. HW and JX assisted in performing the experiments. RW and HL analyzed the data and wrote the paper. RW, HL, GY, and XH revised the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Experiment and verification in the hospital.

[MP4 File (MP4 Video), 34453 KB - humanfactors_v10i1e42870_app1.mp4 ]

References


Abbreviations

CNN: convolutional neural network
EICU: emergency intensive care unit
FAHZU: First Affiliated Hospital, Zhejiang University
GUI: graphical user interface
HCPS: human-cyber physical system
IIT: investigator-initiated trial
IMU: inertial measurement unit
MAR: medical assistive robot
PN2: Perception Neuron 2
ROS: robot operating system
WebRTC: web real-time communications
©Ruohan Wang, Honghao Lv, Zhangli Lu, Xiaoyan Huang, Haiteng Wu, Junjie Xiong, Geng Yang. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 20.04.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Eliciting Opinions on Health Messaging During the COVID-19 Pandemic: Qualitative Survey Study

Sienna Ruiz¹, BA; Uzoma Charles Okere¹, MPH, MD; Michelle Eggers¹, MA; Catina O'Leary², PhD; Mary Politi¹, PhD; Fei Wan¹, PhD; Ashley J Housten¹, OTD, MSCI

¹Division of Public Health Sciences, Department of Surgery, Washington University in St. Louis, St. Louis, MO, United States
²Health Literacy Media, St. Louis, MO, United States

Corresponding Author:
Ashley J Housten, OTD, MSCI
Division of Public Health Sciences
Department of Surgery
Washington University in St. Louis
660 S Euclid Ave, MSC 8100-0094-02
St. Louis, MO, 63110
United States
Phone: 1 (314) 454 7958
Email: ahousten@wustl.edu

Abstract

Background: Effective public health messaging has been necessary throughout the COVID-19 pandemic, but stakeholders have struggled to communicate critical information to the public, especially in different types of locations such as urban and rural areas.

Objective: This study aimed to identify opportunities to improve COVID-19 messages for community distribution in rural and urban settings and to summarize the findings to inform future messaging.

Methods: We purposively sampled by region (urban or rural) and participant type (general public or health care professional) to survey participants about their opinions on 4 COVID-19 health messages. We designed open-ended survey questions and analyzed the data using pragmatic health equity implementation science approaches. Following the qualitative analysis of the survey responses, we designed refined COVID-19 messages incorporating participant feedback and redistributed them via a short survey.

Results: In total, 67 participants consented and enrolled: 31 (46%) community participants from the rural Southeast Missouri Bootheel, 27 (40%) community participants from urban St Louis, and 9 (13%) health care professionals from St Louis. Overall, we found no qualitative differences between the responses of our urban and rural samples to the open-ended questions. Participants across groups wanted familiar COVID-19 protocols, personal choice in COVID-19 preventive behaviors, and clear source information. Health care professionals contextualized their suggestions within the specific needs of their patients. All groups suggested practices consistent with health-literate communications. We reached 83% (54/65) of the participants for message redistribution, and most had overwhelmingly positive responses to the refined messages.

Conclusions: We suggest convenient methods for community involvement in the creation of health messages by using a brief web-based survey. We identified areas of improvement for future health messaging, such as reaffirming the preventive practices advertised early in a crisis, framing messages such that they allow for personal choice of preventive behavior, highlighting well-known source information, using plain language, and crafting messages that are applicable to the readers’ circumstances.

(JMIR Hum Factors 2023;10:e39697) doi:10.2196/39697

KEYWORDS
COVID-19; health messaging; rural populations; urban populations; communication; health information; messaging; dissemination; health equity; prevention; implementation
Introduction

Background

Since the emergence of SARS-CoV-2 in 2019 and its resultant disease, COVID-19, public health communication has rapidly adapted to constantly changing information. Adding complexity to public health messaging, the arrival of variant strains, vaccinations [1], and regional differences in the timing and intensity of disease spread [2] have shifted the course of the pandemic. Given the rapid developments in public health practices, scientific innovations, and epidemiological trends, effective health messaging remains critical for improving public awareness and informing health protocols [3].

In 2002, the US Centers for Disease Control and Prevention (CDC) published manuals such as the Crises and Emergency Risk Communication [4]. The Crises and Emergency Risk Communication manual advocates for trusted sources to be first, be right, be credible, support action, and show respect and empathy toward its audience [4]. However, these principles were not fully applied in the United States in response to the onset of the COVID-19 pandemic. Government and health officials were often not the first to speak on COVID-19, leading the public to question information sources. Limited efforts were made to adapt information to evolving circumstances, and complex concepts such as the risks related to COVID-19 were difficult to convey [4].

In addition, rural populations reported distinct beliefs about the COVID-19 pandemic [5], were overall less likely to engage in COVID-19 preventive health behaviors [6,7], and responded differently from urban populations to specific dissemination strategies for health promotion [8]. Specifically, researchers have found that rural populations may be exposed to various structural barriers (eg, fewer educational opportunities [9]) and express political differences (eg, higher beliefs in individualism [10]) that contribute to them having higher levels of distrust related to preventive behaviors such as vaccination and masking [10,11] than urban populations. Such differences persist because rural communities have experienced more severe impacts of the COVID-19 pandemic than urban communities owing to increased rates of poverty, comorbidities, and low access to health resources [9]. To equitably direct health resources, including health messaging, an understanding of the underlying individual and social contexts among geographically diverse populations is required [12].

Goals of This Study

To address missteps in health messaging early in the pandemic, researchers and public health professionals must examine the efficacy of health messages and identify best practices. Currently, there continues to be a need for efficient health messaging regarding COVID-19 risks, treatment, prevention, and vaccination [4]. Attributes such as clarity, concision, legibility, attractiveness, realistic guidance, and emotional appeal are essential components of successful COVID-19 health messaging [3,13,14]. In this study, we aimed to identify opportunities to improve COVID-19 messages for community distribution by health officials and summarize the findings to inform future messaging. Findings from this study can improve how stakeholders approach health messaging design in various contexts and inform the dissemination of future health messaging that incorporates perspectives from stakeholders across urban and rural settings.

Methods

Setting

Community participants were recruited from 2 regions of Missouri, Southeast Missouri (the Bootheel) and the St Louis metropolitan region (STL), and a small group of health professionals were recruited from St Louis. In the Bootheel, most care is provided by federally qualified health clinics in the absence of major hospitals [15]. The Bootheel has higher rates of poverty, higher chronic disease burden, and more older people as well as lower educational attainment than other regions of Missouri [16]. Outside the cities of St Louis and Kansas City, counties in the Bootheel have some of the highest number of Black populations in the state [17]. In urban STL, access to health care resources is mediated by racialized segregation, with the majority Black populations in North St Louis facing higher rates of comorbidities, increased poverty, and more limited availability of health care resources than the majority White populations in South St Louis [18] despite the presence of several major health care institutions in the area at large. Both the Bootheel and St Louis have similarly low levels of health literacy [16].

Message Review and Identification

From July 2020 to September 2020, the research team reviewed the existing public health messages to be used in the surveys. Two research team members used a search engine (eg, Google [Google LLC]) and social media (eg, Facebook [Meta Platforms, Inc] and Instagram [Meta Platforms, Inc]) to identify local, state, national, and international COVID-19 public health messages. Following the full team review, we chose 2 messages in each of the following two types: (1) risk presentations and (2) infographics. A total of 4 messages were chosen because they varied in content, format, and imaging, and they were widely used in the media. Only 4 messages were selected to ensure adequate time in the web-based interview to fully explore how participants responded to the health information in 2 messages, along with their preferences associated with the overall content, format, and imaging in the selected messages. The selected messages were focused on prevention protocols and presented COVID-19 risk using various visual communication strategies. Their sources represented a range of experts (eg, the World Health Organization [WHO], the CDC, and Doctor of Medicine groups) and are described in detail in Table 1.

https://humanfactors.jmir.org/2023/1/e39697

Ruiz et al
The survey session lasted an average of 1 hour for each participant. Participants received a US $50 gift card for their time.

In the context of the COVID-19 pandemic, when in-person interviews were not considered appropriate or safe, the research team operationalized a web-based approach to capture participants' opinions. To recruit participants, the research team members broadly distributed a web-based survey link via social media (eg, Facebook, Twitter [Twitter, Inc.], and Craigslist [Craigslist, Inc.]). This survey collected the participants' contact information, including their email addresses, which were then kept within an Institutional Review Board–approved, password-protected database. After a participant completed the survey and was found eligible, the study team contacted them to schedule the full survey evaluating health messages. Public participants were eligible to participate if they were (1) self-reported English speakers, (2) aged 18 to 80 years, and (3) residing in either the Bootheel or St Louis. Our age cutoff for eligibility was 80 years owing to limitations in the feasibility of recruiting older adults remotely during COVID-19 surges, concerns over access to technology among this population, and potentially differing risk reduction recommendations for older adults. Participants were considered residents of the Southeast Missouri Bootheel if they lived in Dunklin, Stoddard, Mississippi, Pemiscot, or New Madrid County. Participants were considered St Louis residents if they lived in St Louis City or County. Health care professionals were eligible to participate if they (1) were self-reported English speakers, (2) resided in

### Ethics Approval

All surveys and interview guides were approved by the Washington University Institutional Review Board (#202010069). All research procedures were approved by the Washington University School of Medicine Institutional Review Board.

### Surveys

Given the potential differences between urban and rural populations, we surveyed populations from 2 distinct regions, urban St Louis, Missouri, and the rural Southeast Missouri Bootheel, to assess preferences for COVID-19 messaging. We used a purposive sample recruitment approach in both the St Louis and Bootheel regions because our research team had preexisting connections with community organizations that could aid recruitment in both areas. Our study design followed the principles of pragmatic health equity implementation science by surveying members of the general public in each region and health care professionals on their message preferences to inform the immediate development of new messages [12,23]. We situated our findings on COVID-19 health messaging within the social and economic contexts that the participants reported during their survey.

We surveyed participants to elicit their opinions on COVID-19, including their preferences for preselected COVID-19 messages. The survey session lasted an average of 1 hour for each
either the Bootheel or St Louis, and (3) self-identified as a health professional (eg, Registered Nurse, Licensed Practical Nurse, Doctor of Medicine, or Doctor of Osteopathic Medicine). Staff reached out directly via email or phone to eligible participants and scheduled a web-based appointment on a Health Insurance Portability and Accountability Act (HIPAA)–compliant Zoom (Zoom Video Communications, Inc) account for their participation in the study. Survey data were collected and managed using the REDCap (Research Electronic Data Capture; Vanderbilt University) system hosted at the Washington University [24,25].

To reduce participant burden and enhance the feasibility of survey completion, the messages were divided into 2 sets and randomly assigned to roughly equal numbers of participants by the research team before each survey. The purpose of random assignment was not to determine differences between message sets but to evaluate participants’ opinions on multiple types of health messages.

The research team conducted the surveys with participants between November 2020 and February 2021 (Multimedia Appendix 1). The survey questions followed pragmatic and health equity guidelines by evaluating the social and economic impacts of the pandemic and eliciting real-time opinions on health messaging with the goal of improving message development later in the study. Questions 15, 16, 33, and 61 on the social and economic impacts of the pandemic and questions 7 and 11 to 14 on the exposure of the participants to COVID-19 elicited potential socioeconomic and health inequities between public samples, inspired by calls to compare health indicators and individuals’ social positions (eg, race and ethnicity, socioeconomic status, and educational attainment) to examine potential health inequities [12]. Halfway through the survey, interviewers shared their screen to show participants their assigned messages. After the participants had thoroughly reviewed the messages, the interviewers asked the participants open-ended questions on their opinions of the messages. These questions are listed as questions 35 to 46 in Multimedia Appendix 1 for public participants and questions 18 to 25 in Multimedia Appendix 2 for health care professionals. These questions aimed to identify participants’ perspectives for “tailored implementation, which builds on real-world experiences to identify the participant-identified priorities to address” [12] for the dissemination of improved health messages. The interviewers took notes that closely summarized the participants’ comments. The survey process was audio recorded, and the recordings were stored on a secure university platform.

Shorter, focused surveys were conducted with health care professionals using the same methods, but the questions were designed to capture the needs of their patient populations (Multimedia Appendix 2).

To analyze the qualitative data, the research team members used inductive thematic analysis [26,27]. Data were analyzed and managed using the NVivo software (version 20; QSR International). The purpose of this study was to evaluate participants’ opinions on health messaging. Therefore, we used inductive thematic analysis to gauge how participants viewed each message and the salient themes they discussed in relation to their preferences for the content, design, distribution, and other aspects of the messages. Team members familiarized themselves with the qualitative data by reading through and annotating the interviewers’ notes of each participant’s responses to open-ended questions. Following a close review of the interviewers’ notes, the research team members created a codebook to guide thematic analysis. Once the codebook was finalized, each interview was independently coded by 1 of 3 coders. Then, a separate coder reviewed each coded interview, and discrepancies were discussed and reconciled by the research team to ensure greater reliability. Team members systematically read through, annotated, and summarized each code to create the thematic findings described in the Results section. To qualitatively compare the themes between each participant group, we identified which themes were most salient for each group by examining the degree to which a theme recurred or was important in the sample (ie, themes were considered important if they were “new and advanced understanding, were useful in addressing real-world problems, or did both”) [28]. If the coders found similar levels of recurrence and importance of the same theme in both samples, they listed the theme as salient to both groups, and they found no qualitative difference in their analysis between the groups in relation to each theme.

We used various methods to ensure qualitative rigor [29], such as holding regular team meetings to create the codebook and checking whether coders applied codes consistently across surveys. In our meetings, we also discussed how our backgrounds (eg, from different academic disciplines), our personal experiences of the pandemic, and residing in rural or urban area shaped our approaches to coding and analysis. It was discussed in depth how most authors’ life experiences in urban areas, and 1 author’s life experiences in a rural area, influenced the research team’s understanding of the similarities and differences between urban and rural regions. We consistently examined our interpretations of the thematic results to limit any potential bias toward or against a type of region or the perpetuation of any stereotypes of urban or rural regions.

**Message Redistribution**

In line with the goal of equity in the dissemination of study results to end users [12], we created new COVID-19 messages based on participants’ responses and redistributed these messages for participant feedback. Following the qualitative analysis, the research team created 3 sets of images incorporating participant feedback between March and June 2021. These messages addressed safer summer activities, postvaccination guidelines, and incentives to get vaccinated and were intended to be distributed during the summer of 2021. After the designs were finalized, we recontacted the participants asking them to complete a short survey in June 2021 gathering feedback on the new messages, including whether the new images incorporated their feedback from their initial surveys. The newly created messages and full survey on message redistribution can be found in Multimedia Appendix 3. This redistribution survey approach introduces a low-resource method for eliciting health equity implementation feedback via brief web-based surveys.
Results

A total of 67 participants completed the study, with 31 (46%) community participants from the Bootheel, 27 (40%) community participants from the St Louis area, and 9 (13%) health care professionals from the St Louis area. Overall, 52% (35/67) of participants reviewed message set 1, and 48% (32/67) of the participants reviewed message set 2.

Participant Characteristics

Table 2 presents the sociodemographic information of the total sample. The mean age of the Bootheel public group was younger than that of the St Louis public group (Bootheel mean age 30.3, SD 10.1 years vs St Louis metro mean age 38.0, SD 13.7 years). The health care professional group’s mean age was 34.9 (SD 7.11) years. Health care professionals were either primary care providers (eg, RNs, physicians, and medical assistants) or community health workers (eg, caregivers, social workers, and mental health program managers). Participants across both samples had similarly high levels of health literacy, incomes, and educational attainment, and most participants identified as White or Black.

In terms of COVID-19 exposure, more participants in the Bootheel knew someone close to them who tested positive for COVID-19 (19/31, 61% compared with 13/27, 48% in STL) or who was hospitalized for COVID-19 (25/31, 81% compared with 14/27, 52% in STL). More participants in the Bootheel responded that they could count on people in their neighborhood to help them (28/31, 90% compared with 16/27, 59% in STL) and go to the store for them if they were sick (25/31, 81% compared with 15/27, 56% in STL). Participants in the Bootheel rated the degree to which the pandemic created financial problems for themselves or their family higher than those in St Louis (Table 3). They were also more worried about not being able to access food or important resources, such as transportation or housing, owing to the pandemic (Table 3).

Table 2. Baseline participant characteristics of the final sample (N=67)a.

<table>
<thead>
<tr>
<th>Region</th>
<th>Public (St Louis metro area; n=27)</th>
<th>Public (Southeast Missouri Bootheel; n=31)</th>
<th>Health care professionals (St Louis metro area; n=9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD; range)</td>
<td>38.0 (13.7; 24-67)</td>
<td>30.3 (10.1; 19-68)</td>
<td>34.9 (7.11; 25-47)</td>
</tr>
<tr>
<td>Gender, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Man</td>
<td>12 (44)</td>
<td>20 (65)</td>
<td>2 (22)</td>
</tr>
<tr>
<td>Woman</td>
<td>15 (56)</td>
<td>10 (32)</td>
<td>7 (78)</td>
</tr>
<tr>
<td>Nonbinary</td>
<td>0 (0)</td>
<td>1 (3)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Race and ethnicity, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>15 (56)</td>
<td>13 (42)</td>
<td>6 (67)</td>
</tr>
<tr>
<td>White</td>
<td>11 (41)</td>
<td>16 (52)</td>
<td>3 (33)</td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>1 (4)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>American Indian or Alaska Native</td>
<td>0 (0)</td>
<td>1 (3)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Asian or Pacific Islander</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Other</td>
<td>0 (0)</td>
<td>1 (3)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Education, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than bachelor’s degree</td>
<td>8 (30)</td>
<td>7 (23)</td>
<td>_b</td>
</tr>
<tr>
<td>Bachelor’s degree or higher</td>
<td>19 (70)</td>
<td>24 (77)</td>
<td>_</td>
</tr>
<tr>
<td>Yearly family income, including all sources (US $)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;15,000</td>
<td>2 (7)</td>
<td>2 (6)</td>
<td>—</td>
</tr>
<tr>
<td>15,000-34,999</td>
<td>2 (7)</td>
<td>0 (0)</td>
<td>—</td>
</tr>
<tr>
<td>35,000-54,999</td>
<td>6 (22)</td>
<td>8 (26)</td>
<td>—</td>
</tr>
<tr>
<td>55,000-74,999</td>
<td>3 (11)</td>
<td>15 (48)</td>
<td>—</td>
</tr>
<tr>
<td>≥$75,000</td>
<td>12 (44)</td>
<td>6 (19)</td>
<td>—</td>
</tr>
<tr>
<td>Prefer not to say</td>
<td>2 (7)</td>
<td>0 (0)</td>
<td>—</td>
</tr>
<tr>
<td>Health literacy, mean (SD; range)</td>
<td>14.5 (2.3; 11.0-19.0)</td>
<td>14.2 (2.4; 10.0-18.0)</td>
<td>—</td>
</tr>
</tbody>
</table>

aTotals were calculated by column.

bWe did not collect education, income, health literacy, social, or economic data from health care professionals.
Table 3. Social and economic impacts of COVID-19 for the public participants.

<table>
<thead>
<tr>
<th>Region</th>
<th>Public (St Louis metro area; n=27)</th>
<th>Public (Southeast Missouri Bootheel; n=31)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Have you ever been diagnosed with COVID-19?, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>8 (30)</td>
<td>26 (84)</td>
</tr>
<tr>
<td>No</td>
<td>18 (67)</td>
<td>5 (16)</td>
</tr>
<tr>
<td>Not sure or do not know</td>
<td>1 (4)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Has anyone close to you tested positive for COVID-19?, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>13 (48)</td>
<td>19 (61)</td>
</tr>
<tr>
<td>No</td>
<td>12 (44)</td>
<td>11 (36)</td>
</tr>
<tr>
<td>Not sure or do not know</td>
<td>2 (7)</td>
<td>1 (3)</td>
</tr>
<tr>
<td>How many people do you know who have had COVID-19?, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>5 (19)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>1</td>
<td>3 (11)</td>
<td>4 (13)</td>
</tr>
<tr>
<td>2-5</td>
<td>13 (48)</td>
<td>18 (58)</td>
</tr>
<tr>
<td>≥6</td>
<td>6 (22)</td>
<td>9 (29)</td>
</tr>
<tr>
<td>Do you know anyone who has been hospitalized for COVID-19?, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>14 (52)</td>
<td>25 (81)</td>
</tr>
<tr>
<td>No</td>
<td>11 (41)</td>
<td>5 (16)</td>
</tr>
<tr>
<td>Not sure or do not know</td>
<td>2 (7)</td>
<td>1 (3)</td>
</tr>
<tr>
<td>Do you know anyone who has died from COVID-19?, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>9 (33)</td>
<td>9 (29)</td>
</tr>
<tr>
<td>No</td>
<td>18 (67)</td>
<td>20 (65)</td>
</tr>
<tr>
<td>Not sure or do not know</td>
<td>0 (0)</td>
<td>2 (7)</td>
</tr>
<tr>
<td>I can count on people in my neighborhood to help me if I am sick, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agree</td>
<td>16 (59)</td>
<td>28 (90)</td>
</tr>
<tr>
<td>Disagree</td>
<td>11 (41)</td>
<td>3 (10)</td>
</tr>
<tr>
<td>My neighbors would go to the store for me if I am sick, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agree</td>
<td>15 (56)</td>
<td>25 (81)</td>
</tr>
<tr>
<td>Disagree</td>
<td>12 (44)</td>
<td>6 (19)</td>
</tr>
<tr>
<td>How worried have you been about not being able to afford or access food because of the COVID-19 outbreak? (on a scale ranging from 1 [not worried at all] to 5 [somewhat worried] to 10 [extremely worried]), mean (SD; range)</td>
<td>3.0 (2.6; 0-9)</td>
<td>5.6 (2.8; 0-9)</td>
</tr>
<tr>
<td>How worried have you been about access to important resources such as transportation or housing due to the COVID-19 outbreak? (on a scale ranging from 1 [not worried at all] to 5 [somewhat worried] to 10 [extremely worried]), mean (SD; range)</td>
<td>3.0 (3.3; 0-9)</td>
<td>5.3 (3; 0-9)</td>
</tr>
</tbody>
</table>

Thematic Findings

Overview

We did not identify any qualitative differences between the participants from the St Louis region and those from the Bootheel in how they responded to the messages or in their suggestions for improving the messages. Common themes for all groups included participants’ preference to see the main COVID-19 protocols in messages, desire for personal choice with regard to COVID-19 preventive behaviors, and suggestions for clear and easily accessible source information. Although health care professionals had responses similar to those of both public samples, they more often named health literacy as a factor that could compound the patient’s perceptions and made suggestions for their specific patient populations. Qualitative results are presented in the subsequent sections with italicized interviewer notes used to summarize participants’ responses to open-ended questions in the survey.
Theme 1: Preference for Main COVID-19 Protocols
Many participants recognized the main COVID-19 protocols as behaviors such as hand washing, maintaining 6 ft of social distancing, and wearing a mask [30]. Most participants wanted to see or expected to see these behaviors represented in the messages. Images that were missing these messages were often viewed as incomplete by the participants. One of the participants shared the following:

Yes something is missing, they should include good ways in wearing a mask, information there that shows where a person wear a mask, not leave nose uncovered, chances of transmitting the virus [Interviewer notes of the response of P54 from STL about message 2A]

Another participant said the following:

What about washing hands, other preventive messages...should be part of every message [Interviewer notes of the response of P14 from STL about message 1B]

Theme 2: Desire for Personal Choice in COVID-19 Behavioral Response
The presentation of risks across various activities appeared to resonate with participants’ interest in personal choice or the freedom to make their own choices regarding their health and safety. One of the participants said the following:

I believe people have the right to make their own choices. This isn’t telling people what to do; it just...tells them about the risk. So if you do everything they recommend, your risk is low, but it allows me to make the decision for myself. [Interviewer notes of the response of P205 from the Bootheel about message 1B]

Similarly, another participant said the following:

I don’t feel like they’re telling you what to do, they’re just giving you guidance on how to avoid certain situations and getting COVID. [Interviewer notes of the response of P267 from the Bootheel about message 2A]

A health care professional commented the following:

I like the spectrum rather than do this and don’t do this; more realistic [because] nothing is zero risk [Interviewer notes of the response of health care professional P156 about message 1B]

Another participant said that they liked that the message “doesn’t feel too preachy” [Interviewer notes of the response of P30 from STL about message 2A].

Theme 3: Clear and Easily Accessible Source Information
Most participants described “good” source information as being apparently authentic because of the presence a large logo, coming from a trusted source, and including resources for follow-up. Follow-up could mean obtaining more information about the message or COVID-19 or receiving contact information on whom to call in case one experiences COVID-19 symptoms. One of the participants said the following:

[It’s missing] maybe the CDC website or something...I don’t know who this is coming from. I should trust this, I guess...it’s missing the CDC or something. [Interviewer notes of the response of P23 from STL about message 1B]

Another participant said that the message should provide “a piece of contact information, such as a number to call...There should be information on who to contact if I suspect someone has COVID-19, is exhibiting symptoms” [Interviewer notes of the response of P192 from the Bootheel about message 1A]. A participant also remarked that the message “had no source, web link...[I am] not likely [to follow-up]. I don’t know [the] journal and don’t see [the source] as a link” [Interviewer notes of the response of P200 from the Bootheel about message 2B].

After we asked them which sources in a provided list they used before, they then identified which source they used the most as a free-response answer. The most preferred sources among the participants in St Louis were local news; social media, such as Twitter and Facebook; the WHO; and the CDC, whereas the most preferred sources in the Bootheel were social media, the WHO, and newspapers. For health care professionals, the most preferred sources were the CDC, newspapers, and local news stations.

Health Care Professional Findings
Health care professionals contextualized their suggestions within the applicability of the messages to their patients. They assessed whether the actions outlined in the messages were applicable to their patient populations with limited health literacy or who were older, had low income, or spoke English as a second language. One of the health care professionals said the following:

I think [telling people to stay home when they’re sick] triggers people. A lot of people...can’t do that because of their financial situation, lack of sick leave, or other things. [Interviewer notes of the response of health care professional P85 about message 1A]

One of the providers gave the following answer:

A lot of it [would be confusing] for my patients, most of my patients speak Spanish. [Interviewer notes of the response of health care professional P156 about message 1A]

Another provider said the following:

For some, not everything in here might...be practical. For example, staying 6 feet apart might not be practical for people...[like for those] sharing an apartment or a house with multiple people. [Interviewer notes of the response of health care professional P264 about message 1A]

A health care professional who worked in a health home answered that the advice regarding avoiding close contact would be hard because “some patients like that physical contact...Some people are also hard of hearing, so you would have to get close
Findings in the Context of Health Communication Best Practices

Community participants’ suggestions for message improvement aligned with the best practices for health literacy [31]. These practices included using clear, easily understandable language; visually prioritizing the most important messages; avoiding extraneous information; sufficiently spacing out images and text; using eye-catching colors; visually representing a diverse set of people; incorporating an emotional appeal; and clearly representing the source of the message. Refer to Table 3 for participants’ quotes.

Although the health care professionals’ suggestions also aligned with the principles of health literacy, they were more likely to specifically reference the terms “literacy” or “health literacy” when gauging the potential impact of the message. For example, one of the health care professionals commented that “some of the visual language is less clear, people with low literacy would be [confused]” (Interviewer notes of the response of health care professional P156 about message 1A). Another health care professional said the following:

*I think it’s highly detailed if you have the time and literacy...but as a general service announcement, I don’t think it’s that effective.* [Interviewer notes of the response of health care professional P246 about message 2B]

Yet another health care provider said the following:

*I think it’s really good but there’s a lot of blocks, which I think someone educated with good eyesight that’s fine, but for someone who is older or low literacy that is too much going on.* [Interviewer notes of the response of health care professional P251 about message 1B]

For more suggestions and quotes on this topic, see Table 4.
Table 4. Public participants’ suggestions for improving the messages with health literacy principles.

<table>
<thead>
<tr>
<th>Participants’ suggestions</th>
<th>Examples and quotes</th>
</tr>
</thead>
</table>
| Use clear language that is easy to understand; vague terms without definitions are confusing. | • Examples of phrases that were confusing:  
  - “Reopen intelligently”  
  - Vague use of “duration”  
  - “When near people, wear a mask”  
  - “Forceful exhalation”  
  - “Face covering”  
  - “High or low occupancy”  
  - “Opening intelligently” |
| Ensure that the “most important” images and messages stand out by making them larger and placing them along the top or top left. | • “The mask is a message that needs to be reinforced. If people are going to look at anything, they’ll look at the top row. The middle is busier, so people won’t glance at that, they’ll glance at the top” (Interviewer notes of the response of health care professional P85 about message 1A). |
| Remove any information that is not strictly necessary to prevent overwhelming viewers. | • “What’s really good about this piece is that it puts so much information in one space there is no unnecessary information and it is clear even for people that may not fully understand English” (Interviewer notes of the response of P224 from the Bootheel about message 2A).  
  • “I think this one is not as good as the other one. I feel like people are not as likely to really decipher through all the color coding and different info. I feel like the other was more straight forward, direct, easy. This one you have to spend a little more time with it and dig into it” (Interviewer notes of the response of P131 from STL about message 2B). |
| Ensure that the image is not busy, cluttered, or cramped, and sufficiently space out text and images. | • “I feel like it’s too much. They could make it simpler. I can’t even read it, the print is too small. I would need glasses. For example, if this was hung up in a restaurant, I wouldn’t stop to look at it cause it’s just too much, and the print is too small” (Interviewer notes of the response of P269 from the Bootheel about message 1B).  
  • “Too info dense; too much wording...given the format it’s cluttered and crowded with too much text” (Interviewer notes of the response of P156, from STL about message 1B). |
| Colors chosen for the image should enhance the attractiveness and understandability of the message. | • “It is a lot more clear because of the colors; [I] suggest a lot more colors and brighter colors so it is more eye-catching” (Interviewer notes of the response of P268 from the Bootheel about message 1B).  
  • “It is beautiful for the color which makes it easier to understand” (Interviewer notes of the response of P219 from the Bootheel about message 2B).  
  • “It catches your attention, the bright colors draw you in” (Interviewer notes of the response of P15 from STL about message 2A). |
| People in the images should be diverse (eg, gender, race, and ethnicity) but more realistic looking. | • “Better images—use real individuals to be more legible, not every person can like cartoons, real people be better” (Interviewer notes of the response of P146 from STL about message 1A).  
  • “I think I’d prefer eyes, nose, and mouth on people. It does look a little funny. I like the diversity of it” (Interviewer notes of the response of P23 from STL about message 1A). |
| Messages should have emotional appeal to be effective. | • “Message like this could appeal more to people’s human nature, something to suggest this is dangerous, people are dying and this is very important, this is informative but doesn’t touch people’s emotions” (Interviewer notes of the response of P05 from STL about message 1A). |

Message Redistribution

On the basis of the survey feedback on our first message sets, we designed new messages to reflect participants’ perspectives. Specifically, we used a list of clear questions rather than directives so that messages could be more readily received and allow readers to make various choices regarding preventive behaviors. We also depicted a diverse (eg, race and ethnicity and age) range of people and activities (eg, eating and outdoor activities) and provided a section on masks that reinforced the main COVID-19 protocols and a link for learning more to establish greater trust with the source. Using the same principles, we also created a message set dedicated to clarifying the main COVID-19 protocols and a link for learning more to establish greater trust with the source. Our third message set used distinct colors and clear, simple imagery to showcase positive reinforcements for getting vaccinated.

Of the original 65 participants we were able to reach via email (2 participants did not provide an email or gave invalid email addresses), 54 completed the survey, leading to an 83% completion rate. Most participants had an overwhelmingly positive response to the new messages and agreed that the new messages incorporated their feedback from the surveys. Overall, the participants liked the content, bright colors, and simple wording. Common themes expressed by most participants were that they appreciated the simple, precise wording and liked the bright, distinct colors that caught readers’ attention and positive emotional appeal. A participant in St Louis (P23) said that the reminders of what people could do after vaccination “shines”
(Interviewer notes of the response of P23 from STL). One of the participants commented the following:

“I’m quite impressed by how simple and illustrative the messages are and by just a quick glance I’m able to understand what message the sender wants to portray.” [Interviewer notes of the response of P43 from the Bootheel]

Another participant said the following:

“Yes, [they included my feedback], most certainly so. They made the words larger so everyone can see and also they used more graphic pictures that can be interpreted easily.” [Interviewer notes of the response of P225 from the Bootheel in response to the second question in the follow-up survey; full follow-up survey in Multimedia Appendix 3]

Discussion

Principal Findings

Despite the anticipated differences between the urban and rural populations’ responses to COVID-19 health messages, both groups responded similarly. Both wanted health messages that were consistent, were attractive, were accessible, and emphasized choice in behavioral responses to the pandemic. Furthermore, although our public sample in the Bootheel may have experienced higher COVID-19 exposure and worse social and economic impacts of the pandemic, as indicated by their response to our survey questions on COVID-19, and thus could have had more particular desires for messages owing to personal contexts, the messaging preferences were largely the same between the Bootheel and St Louis samples. This result differs from studies that have found differences between urban and rural populations’ responses to COVID-19 messages [32] and other health messaging campaigns [33]. This may be because both samples had similarly higher levels of health literacy, income, and educational attainment, which may support the participants from both samples to more critically analyze and apply health messages than those with limited health literacy, lower incomes, and lower educational attainment [34]. However, our detailed findings related to people’s similar preferences for the display and content of health messaging might suggest that the socioeconomic, cultural, and political differences between urban and rural communities [8] should not overshadow the development of broadly applicable and well-designed messaging. Although health officials should consider using unique communication channels to reach rural residents [8], such as local newscasters or community health care professionals trusted by the participants in our study, regional differences should not obfuscate the creation of well-designed health messages at the state or national level.

The participants described their preference for COVID-19 protocols to be succinctly presented in each message they saw. They were especially drawn to messaging that called for the protocols to be succinctly presented in each message they saw. Although health officials should consider using unique communication channels to reach rural residents [8], such as local newscasters or community health care professionals trusted by the participants in our study, regional differences should not obfuscate the creation of well-designed health messages at the state or national level.

The trustworthiness and accessibility of the source of information generated concern among the participants. They wanted COVID-19 messaging to present the possible repercussions of nonadherence to protocols to inform individuals’ decisions. The importance of personal choice may reflect American beliefs surrounding individual liberties, and messages that appear to infringe on personal freedoms can lead to a decreased likelihood of enacting preventive behaviors [36,37]. Positive resonance with such symbols of health interventions could influence people’s reception of the messaging itself. More research is needed to examine people’s relationships with basic preventive health behaviors to help create messaging that can reassure the public and help encourage adherence to such behaviors during periods of uncertainty or rapidly changing safety recommendations [38].

The participants also preferred that personal choice be reflected in COVID-19 messages. That is, they wanted COVID-19 messaging to present the possible repercussions of nonadherence to protocols to inform individuals’ decisions. The importance of personal choice may reflect American beliefs surrounding individual liberties, and messages that appear to infringe on personal freedoms can lead to a decreased likelihood of enacting preventive behaviors [36,37]. Positive resonance with such symbols of health interventions could influence people’s reception of the messaging itself. More research is needed to examine people’s relationships with basic preventive health behaviors to help create messaging that can reassure the public and help encourage adherence to such behaviors during periods of uncertainty or rapidly changing safety recommendations [38].

The trustworthiness and accessibility of the source of information generated concern among the participants. They wanted to see credible sources and suggested including larger logos for trusted sources, such as the WHO or CDC. They also wanted to see contact information for sources, such as phone numbers or websites. Participants across our samples listed local sources such as friends, family members, local news, physicians, or other health care professionals as their most used sources of information on COVID-19. This finding is consistent with studies that found that facilitating relationships with local stakeholders and health care providers is essential for building trust in COVID-19, especially in rural communities [42,43]. To increase people’s trust in message sources, we recommend including contact information and a specific link to learn more about the health issue as well as using a knowledgeable

https://humanfactors.jmir.org/2023/1/e39697
Web-based methods have the benefit of being more accessible and less resource intensive and time consuming than other research methods [54].

Informed by our findings, we created refined health messaging that incorporated the themes participants discussed during their surveys to disseminate examples of health messaging that both incorporated participants’ varied preferences and aligned with health communication best practices. These messages were action oriented and uniquely addressed personal choice in health prevention, common health protocols, and accessible source information. We used a list of questions to prompt readers to consider their risk when planning activities to present less overwhelming visual content and align with participants’ preferences for personal choice. When communicating complex topics, such as personal risk and probabilities, researchers and public health officials often use visual depictions, such as icon arrays and figures, to help enhance the understanding of numerical estimates [55]. However, high amount of numerical information has the potential to overwhelm viewers, especially those with limited overall literacy or health literacy [56-58]. Future risk messaging might consider using gist representations of risk to inspire readers to consider the general magnitude of their risk [55]. For those seeking more precise, verbatim risk information, links or QR codes can provide more detailed probabilistic information. Incorporating numeric information that is easy to understand can guide the development of engaging and useful health messages.

Limitations, Strengths, and Future Directions

This study has multiple limitations and strengths that indicate potential avenues for future research on people’s opinions related to health messaging. First, we used self-selection methods for recruitment, which may have attracted individuals who were highly motivated to participate in a study related to COVID-19. These methods may have also resulted in samples of people with higher incomes, educational attainment, and health literacy scores than the general public in both St Louis and the Bootheel. At the time of data collection, which was during the early phases of the COVID-19 pandemic, remote recruitment based on self-selection was our only recruitment option, which likely limited the populations we were able to reach for our study. These methods may also have resulted in samples of people with higher health literacy scores, incomes, and educational attainment than the general public in both St Louis and the Bootheel. Such selection bias may suggest that participants were more predisposed to respond positively to COVID-19 mitigation efforts and express preferences for health literacy principles and benefits many populations in the United States, including racial or ethnic minorities, groups with lower educational attainment, and those with low socioeconomic status [48,49]. We suggest that designers familiarize themselves with the principles of health literacy [50] and incorporate them into the development of future health messaging. Health literate approaches include using plain language to be concise and conversational [46] and incorporating prosocial messages that emotionally compel readers to comply [51].

Following our initial analysis, the results of which indicated similarities in messaging preferences, we created a short web-based survey for original participants to comment on new messages created based on their original surveys, continuing participants’ engagement in the research process. Our high completion rate for the survey on message redistribution and participants’ appreciation for the incorporation of their feedback from the initial surveys indicate the importance of continued contact with research participants. Disseminating results back to participants and engaging them throughout the message development process can improve the trust in researchers and strengthen the ties between research organizations and various communities. Other studies have found that creative methods of recontacting participants and disseminating results in the form of community listening sessions or research forums can improve the willingness to participate in research [52,53]. Building on this literature, short web-based surveys and community-based message creation may add to the data collection methods that health literacy researchers can use when attempting to engage participants in the research process. Such web-based methods have the benefit of being more accessible and less resource intensive and time consuming than other research methods [54].
studies that document that even though rural populations are less likely to participate in preventive health behaviors related to COVID-19, they may still highly believe in the efficacy of public health measures and the threat of the pandemic to their community and be open to receiving health messages from trusted local health officials [59,60]. Our findings may also indicate that it is important that researchers not homogenize rural populations’ approaches to the pandemic and instead dedicate more resources to addressing how rural populations understand their health. Furthermore, although we did not evaluate participants’ level of understanding of the health messages, we know that the mastery of what people attend to in health messages is vital in how we design and distribute health messages and inform the public. Future research can evaluate whether disseminating appealing public health messages translates into the comprehension of the message content. Moreover, we recruited participants from a Midwestern state in the United States, meaning that the results may not be applicable to other geographic areas, and our samples did not include racial and ethnic groups that were not White or Black. However, our mixed methods approach and thematic analysis revealed areas of improvement that can strengthen public health messaging and reinforce the importance of best practices for effective health messaging.

In addition, although our data represent participant perspectives from a relatively early point in the pandemic, the message redistribution method may continue to prove useful when examining other health literacy issues in the context of urban and rural health disparities. These disparities continue to be observed in cancer prevalence [61], cardiovascular care [62], and other health domains. More research is required to fully examine local contexts and attitudes toward COVID-19 messaging, but our findings can improve and inform public health messaging so that it is as clear, applicable, and effective as possible.

Conclusions
This analysis of participants’ responses indicates areas of improvement for future health messaging, such as reaffirming common COVID-19 protocols, framing content such that it allows for personal choice, and advertising easily accessible source information. Messages communicated by health care professionals should align with the needs of specific patient populations, and all messages must include plain language, effective wording, emotional appeal, and an attractive design. Participants’ engagement in message creation can aid in health equity implementation. These findings are critical for stakeholders developing public health messages for the COVID-19 pandemic and other public health crises.

Acknowledgments
Research reported in this publication was supported, in part, by the Washington University Institute of Clinical and Translational Sciences grant UL1TR002345 from the National Center for Advancing Translational Sciences (NCATS) of the National Institutes of Health (NIH). The content is solely the responsibility of the authors and does not necessarily represent the official view of the NIH.

Data Availability
Data are available upon reasonable request to the corresponding author. Data are not publicly available to protect the confidentiality of the study participants.

Conflicts of Interest
MP was a consultant for UCB biopharma in 2022 on a topic unrelated to this manuscript.

Multimedia Appendix 1
The survey given to the members of the public samples during this study.
[DOCX File, 39 KB - humanfactors_v10i1e39697_app1.docx]

Multimedia Appendix 2
The survey used for the providers in our study.
[DOCX File, 20 KB - humanfactors_v10i1e39697_app2.docx]

Multimedia Appendix 3
The messages we created in response to participant feedback as well as the brief survey we used to collect participants’ opinions on the newly created messages.
[DOCX File, 893 KB - humanfactors_v10i1e39697_app3.docx]

References


46. Plain language materials and resources. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/healthliteracy/developmaterials/plainlanguage.html [accessed 2021-08-18]


50. Parker R, Jacobson K. Health literacy principles: guidance for making information understandable, useful, and navigable. NAM Perspectives 2014 Dec 22;4(12) [FREE Full text] [doi: 10.31478/201412g]


cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
The Impact of Individuals’ Social Environments on Contact Tracing App Use: Survey Study

Atiyeh Sadeghi1, PhD; Sebastian Pape1, PD, PhD; David Harborth1, PhD
Chair of Mobile Business & Multilateral Security, Faculty of Economics and Business, Goethe University Frankfurt, Frankfurt, Germany
*all authors contributed equally

Corresponding Author:
Sebastian Pape, PD, PhD
Chair of Mobile Business & Multilateral Security
Faculty of Economics and Business
Goethe University Frankfurt
Theodor-W.-Adorno-Platz 4
Frankfurt, 60323
Germany
Phone: 49 69798 ext 34701
Email: sebastian.pape@m-chair.de

Abstract

Background: The German Corona-Warn-App (CWA) is a contact tracing app to mitigate the spread of SARS-CoV-2. As of today, it has been downloaded approximately 45 million times.

Objective: This study aims to investigate the influence of (non)users’ social environments on the usage of the CWA during 2 periods with relatively lower death rates and higher death rates caused by SARS-CoV-2.

Methods: We conducted a longitudinal survey study in Germany with 833 participants in 2 waves to investigate how participants perceive their peer groups’ opinion about making use of the German CWA to mitigate the risk of SARS-CoV-2. In addition, we asked whether this perceived opinion, in turn, influences the participants with respect to their own decision to use the CWA. We analyzed these questions with generalized estimating equations. Further, 2 related sample tests were performed to test for differences between users of the CWA and nonusers and between the 2 points in time (wave 1 with the highest death rates observable during the pandemic in Germany versus wave 2 with significantly lower death rates).

Results: Participants perceived that peer groups have a positive opinion toward using the CWA, with more positive opinions by the media, family doctors, politicians, and virologists/Robert Koch Institute and a lower, only slightly negative opinion originating from social media. Users of the CWA perceived their peer groups’ opinions about using the app as more positive than nonusers do. Furthermore, the perceived positive opinion of the media (P=.001) and politicians (P<.001) was significantly lower in wave 2 compared with that in wave 1. The perceived opinion of friends and family (P<.001) as well as their perceived influence (P=.02) among nonusers toward using the CWA was significantly higher in the latter period compared with that in wave 1. The influence of virologists (in Germany primarily communicated via the Robert Koch Institute) had the highest positive effect on using the CWA (B=0.363, P<.001). We only found 1 decreasing effect of the influence of politicians (B=−0.098, P=.04).

Conclusions: Opinions of peer groups play an important role when it comes to the adoption of the CWA. Our results show that the influence of virologists/Robert Koch Institute and family/friends exerts the strongest effect on participants’ decisions to use the CWA while politicians had a slightly negative influence. Our results also indicate that it is crucial to accompany the introduction of such a contact tracing app with explanations and a media campaign to support its adoption that is backed up by political decision makers and subject matter experts.

(JMIR Hum Factors 2023;10:e45825) doi:10.2196/45825

KEYWORDS
contact tracing app; corona warning app; Corona-Warn-App; social influence; usage; COVID-19
Introduction

Background

With the global pandemic caused by SARS-CoV-2, digital proximity tracing systems to identify people who have been in contact with an infected person are one approach to trying to get the pandemic under control. There have been many discussions on different implementations and their architecture [1], that is, whether the approach should be centralized or decentralized. One implementation is the German Corona-Warn-App (CWA). It is built with privacy in mind, is based on a decentralized approach [2], and the usage intention of German citizens has already been widely discussed concerning privacy concerns [3] and knowledge about the app [4]. However, the influence of different groups in the social environments of citizens on the use of contact tracing apps during the pandemic was—to the best of our knowledge—not a subject of extensive research before. This is interesting from a theoretical point of view because research on the acceptance of new technologies considers social influence as an antecedent of behavioral intention to use technologies [5]. Consequently, it also found its way [6,7] into some successors of the Technology Acceptance Model (TAM; cf. [8]).

Furthermore, this lack of research on social influence and contact tracing apps is surprising because the medical nature of the disease (SARS-CoV-2) is inherently based on human interactions. Furthermore, previous research suggests that knowledge about the CWA significantly reduces the privacy concerns about it [4]. However, most citizens do not acquire knowledge from primary sources but rather from discussions with their peer groups. Thus, the assumption must be made that the decision to undertake a disease prevention measure (in our case using a contact tracing app) is always embedded within the back and forth of social interactions, perceptions, or even pressures. This can also be seen in the design of contact tracing apps. They not only allow their users to see whether they had potential contact with infected individuals but also to warn others by entering positive (or negative) SARS-CoV-2 test results. Consequently, it is crucial to investigate how citizens perceive the opinion of their peer groups on using contact tracing apps. However, because this question alone would not suffice to draw conclusions on the decision of the citizens to use the app, we also need to ask whether this opinion influences them for or against using such an app. To address this, we conducted a longitudinal survey study with 833 participants to investigate these opinions and the perceived influence of a set of peer groups on the participants to use the CWA. Peer groups in our study include media (eg, print media, websites, and television), family doctors, politicians, virologists/the Robert Koch Institute (RKI; a German federal government agency and research institute responsible for disease control and prevention), social media, and friends and family.

We surveyed participants 2 times with a time distance between the surveys of approximately 10 months to also investigate changes over time of the use behavior of the app and the opinions and influences of the relevant groups and to control for the severity of the pandemic. These 2 periods were chosen because we observed the height of the death rates due to SARS-CoV-2 in Germany during the first period (wave 1) with more than 1200 deaths at a given day compared with significantly lower death rates during the second period (wave 2) with approximately 200 deaths at a given day.

In summary, we investigate the following 4 research questions (RQs):

- **RQ1:** How do users and nonusers perceive opinions of relevant groups and their influence?
- **RQ2:** What are the differences between users and nonusers?
- **RQ3:** How do the opinions and the influence change over time (from wave 1 to wave 2) driven by infection rates (decreased from wave 1 to wave 2)?
- **RQ4:** How does the opinion of the relevant groups influence the usage of the CWA?

Prior Work

Researchers have conducted surveys on adopting SARS-CoV-2 tracing apps in various countries [9]. Although some data point to reasonably high app support globally [10], other research highlighted the issue of low usage rates [11]. The majority of articles use surveys to investigate the users’ adoption of 1 or more contact tracing apps (eg, in Australia [12], China [13], France [10], Germany [3,4,10,13,14], Ireland [15,16], Italy [10], Taiwan [17], the United Kingdom [10,18,19], and the United States [10,13,20]). For example, Horstmann et al [21] (see also [3]) found for a sample in Germany that the most common reasons for nonusers were privacy concerns, lack of technical equipment, and doubts about the app’s efficiency. Most other studies reported similar results and identified privacy concerns as one of the main barriers to using contact tracing apps. In particular, people are worried about corporate or government surveillance, potentially even after the pandemic [16], leakage of data to third parties [10], exposure of social interactions [22], and secondary use of the provided data [22]. However, misconceptions based on widespread knowledge gaps accompany the adoption of contact tracing apps [4].

Besides these studies, Blom et al [23] studied potential adoption barriers of the official contact tracing app (Corona-Warn-App) that was launched in Germany on June 16, 2020.

Their findings indicate that with low adoption rates in the general population and problems with selectivity across subgroups, the data reflect a pessimistic view of the usefulness of app-based contact tracing to contain the SARS-CoV-2 epidemic in Germany. According to their estimates, roughly 81% of the German population aged between 18 and 77 years have access to devices that can be used to install the German Corona-Warn-App. However, the authors found that only 35% are eager to do so. This indicates that most citizens lack awareness about the app or the motivation to use it. Thus, research is needed to investigate individuals’ reasons for and against using the app.

Previous studies have focused on users’ perceptions and motivations concerning mobile health (mHealth) apps on a more general level without considering the aspect of social interactions and pressure, which are associated with a technology focusing on combating infectious diseases [24-26]. According
to prior research on individuals’ motivations for using mHealth apps, factors such as access to a smartphone with the necessary app installed and internet connectivity [27,28], smartphone users’ capacity to carry out the functions necessary to use the app [29], prior experience using mobile technologies [30,31], reliable information and true performance and functionality provided by the apps [32-35], trust in data security or authorities [10,14,36,37], and privacy concerns [16,24,38-51] have a significant role in their motivation to use mHealth apps.

Less research has, however, examined the effect of social influence and social relationships [10,14,16,52-59] on the motivation to use mHealth apps, especially in the context of infectious disease presentation (which effectively is the target of contact tracing apps). For mHealth apps in general, research finds that the more people identify with others, the more positively they view these other individuals [60-62]. The degree of identification with the source (or “authority”) predicts the propensity of individuals to utilize these new technologies [63,64]. Social influence is also used in related research, which uses the TAM to investigate factors influencing users’ willingness to use and pay for a mobile health care app [59]. Bettiga et al [59] incorporated the idea of social influence through subjective norms that play a crucial part in decisions and health-related choices. A subjective norm is defined as an individual’s sense of the level to which significant others approve or disapprove of the target behavior [65]. Self-care and preventative behavior are frequently driven by a sense of compliance to social expectations from family members, the social group to which the individual belongs, and doctors. This research also shows that the general intention to accept preventative mHealth technology is influenced by the social influence of healthy adults. In another investigation, people used social interactions with their peers as an active information-seeking strategy to rule out potential negative effects of using or not using a certain technology. In this way, social interaction assists in lessening uncertainty by serving as a mechanism for gathering knowledge and excluding alternatives [14].

Li et al [66] evaluated a model of trusting bases along with 8 different factors in the context of initial trust in a national identity system. They found that in the setting of initial trust, social influence had a greater impact on trusting beliefs than any of the trusting bases. It is crucial because initial trust formation is particularly pertinent in information systems, where users must get past their concerns about risk and uncertainty before utilizing a technology. The closest related work to ours is the one by Scholl and Sassenberg [52], which explored whether a person’s level of identification with 2 groups, namely, (1) with the beneficiaries of app use (ie, people in their social surroundings) and (2) the source endorsing the app (ie, government officials) predicts their propensity to utilize contact tracing apps. Their results indicate that the more people identify with members of their social environment (the beneficiaries) and the government (the source), the more their app acceptance increases. We have focused on the opinion of more groups with the lens of social influence as a key driver due to the context of using the app to prevent an infection with an infectious disease and warn other members of the society in case one is sick. Therefore, we contribute to the literature by increasing the detail concerning the specific social group in question and disentangling potential relations among the influencing powers of these different groups.

Methods

Overview

In this section, we briefly cover the data collection, sample demographics, and the questionnaire development (see Multimedia Appendix 1 for the questionnaire).

Data Collection and Demographics

We conducted the study with a certified panel provider in Germany (certified following the ISO 20252 norm [67]). The survey was implemented with the software LimeSurvey (version 2.72.6; LimeSurvey GmbH) [68], hosted on a university server and conducted in 2 waves. The first wave was ran in January 2021 and the second wave was ran from mid-October 2021 to mid-November 2021.

The idea behind the 2 waves was to collect data in 2 points of time with different acuteness and severeness of the pandemic (Figures 1 and 2). We chose hospitalization and death rates, as politicians in Germany decided upon disease prevention measures (eg, lockdowns) based on these 2 measures later in the course of the pandemic (initially, the incidence rate was used as the main indicator for political decisions).

In the first wave, we sampled the participants to achieve a representative sample for Germany. For that purpose, we set quotas to end up with approximately 418/833 (50.2%) females and 415/833 (49.8%) males in the sample and distribution of age following the EUROSTAT 2018 census [69]. Furthermore, we set a quota to end up with half of the sample using the CWA and the other half not using it.

In the second wave, we could only rely on the participants of the first wave. Therefore, we did not sample using hard quotas but steered participation by sending out invitations to participate in bunches. Each bunch addressed the underrepresented participants to balance the properties use of the CWA, age, and gender.
**Questionnaire**

To assess the opinion of relevant peer groups and their influence on the participant, we asked 2 questions in a matrix, where the participant was asked about each peer group’s opinions on the app’s usage as well as how the opinion of each group influenced the participant for or against using the CWA. There was no suitable construct, thus we developed the 2 questions based on existing literature on perceived opinions [7] and influence [8] of related research. As relevant peer groups, we identified media (e.g., print media, websites, television), family doctors, politicians, virologists/RKI (a German federal government agency and research institute responsible for disease control and prevention), social media, and friends/family based on discussions in the public press. The items for the peer groups’ opinions were measured with a 7-point Likert scale, ranging from “1=strongly negative” to “7=strongly positive.” The items for the peer groups’ influence were measured with a 7-point Likert scale, ranging from “1=strongly against the use of the app” to “7=strongly for the use of the app.” In addition, we gathered the demographics age, gender, education, and income of the participants.

We conducted a pretest with 12 researchers in a workshop. Each researcher answered the question independently. Afterward, we discussed the items and clarified their understanding and meaning. For perceived opinion and influence, only minor changes were made concerning the peer group names.

**Ethical Considerations**

Users were informed about the purpose of the study, about the storage location of the survey data, and that they stay anonymous.
as long as they do not reveal their identity within the free texts. However, we used an identifier from the panel provider to link the date for each participant across the 2 waves. We did not have any further information from the panel provider linked to the identifier. Minors were not allowed to participate. This was ensured by our panel provider and an additional information text before our survey. Participants agreed that their data are used for research and consequent publications.

The user study was evaluated by the Joint Ethics Committee of the Faculty of Economics and Business of Goethe University Frankfurt and the Gutenberg School of Management and Economics of the Faculty of Law, Management and Economics of Johannes Gutenberg University Mainz. The project has been classified as “ethically acceptable.”

Data Analysis
The data have been analyzed using SPSS version 26 (IBM, Inc.) and R (R Foundation). In the first step, descriptive statistics were used to show how users and nonusers perceived the opinions of relevant groups and their influence. In the second step, as the data were not normally distributed, 2 related samples tests (including mean, SD, minimum, maximum, number of nonmissing cases, and quartiles. Tests: Wilcoxon signed-rank, sign, McNemar) and nonparametric tests (Wilcoxon) were applied to understand how the opinions and the influence differed between users and nonusers and changed over time (from wave 1 to wave 2). And finally, using the marginal model with the generalized estimating equations, we estimated how the different groups in the participants’ social environments influenced the usage of the CWA.

Results
Overview
In this section, the result of the data analysis is reported. We have 2 main parts in this section: First, we briefly discuss RQ1, which is primarily a descriptive analysis of our sample. Then, second, in the data analysis part, we present the results of the remaining 3 RQs.

Data Collection and Demographics
Our sample from the first wave consisted of 1752 participants. Following EUROSTAT 2018, participants were representatives of Germany concerning age and gender, income, and education (cf. [3,4]); 896 participants use the CWA (51.14%), whereas 856 do not (48.86%). As this is a longitudinal study with the goal to compare changes over time, we only considered the participants that took part in waves 1 and 2. This left us with 833 participants who were roughly split into 2 equally sized groups of the CWA users and nonusers (Table 1).

As we deliberately divided the sample into 2 approximately equal groups (CWA users and nonusers), we needed to ensure that the groups were not biased with respect to the demographics (Table 2). For age, we conducted a Shapiro-Wilk test for normality and found that the variable was not normally distributed ($P<.001$). Therefore, we used a Wilcoxon signed-rank sum test and found that there were no significant differences in terms of age between CWA users and nonusers ($P=.85$). We also conducted Pearson chi-square tests and found that age ($P=.62$) and gender ($P=.09$) did not reveal a statistically significant difference between users and nonusers. However, for income ($P=.002$) and education ($P=.008$), there were statistically significant differences between users and nonusers, with both of these variables being statistically significantly higher for the users compared with the nonusers. To evaluate the effect size, we additionally conducted Kendall $\tau$ test and found that the correlation between users/nonusers and their income ($P=.01$, $\tau=0.085$) as well as education ($P<.001$, $\tau=0.116$), respectively, was only small. Based on this result, we argue that the absolute difference does not have a substantial confounding effect on our later analysis.

### Table 1. Participant’s use of the Corona-Warn-App over time.

<table>
<thead>
<tr>
<th>Usage/wave</th>
<th>Wave 1 (N=833)</th>
<th>Wave 2 (N=833)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Users</td>
<td>409</td>
<td>427</td>
</tr>
<tr>
<td>Nonusers</td>
<td>424</td>
<td>406</td>
</tr>
</tbody>
</table>
Table 2. Demographics of participants who took part in both waves (N=833).

<table>
<thead>
<tr>
<th>Demographics</th>
<th>Value, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td></td>
</tr>
<tr>
<td>18-29 years</td>
<td>118 (14.2)</td>
</tr>
<tr>
<td>30-39 years</td>
<td>149 (17.9)</td>
</tr>
<tr>
<td>40-49 years</td>
<td>166 (19.9)</td>
</tr>
<tr>
<td>50-59 years</td>
<td>214 (25.7)</td>
</tr>
<tr>
<td>60 years and older</td>
<td>186 (22.3)</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>418 (50.2)</td>
</tr>
<tr>
<td>Males</td>
<td>415 (49.8)</td>
</tr>
<tr>
<td>Divers</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Prefer not to say</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Net income</strong></td>
<td></td>
</tr>
<tr>
<td>€500–€1000</td>
<td>76 (9.1)</td>
</tr>
<tr>
<td>€1001–€2000</td>
<td>177 (21.2)</td>
</tr>
<tr>
<td>€2001–€3000</td>
<td>202 (24.2)</td>
</tr>
<tr>
<td>€3001–€4000</td>
<td>146 (17.5)</td>
</tr>
<tr>
<td>More than €4000</td>
<td>156 (18.7)</td>
</tr>
<tr>
<td>Prefer not to say</td>
<td>76 (9.1)</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
</tr>
<tr>
<td>No degree</td>
<td>3 (0.4)</td>
</tr>
<tr>
<td>Secondary school</td>
<td>99 (11.9)</td>
</tr>
<tr>
<td>Secondary school b</td>
<td>278 (33.4)</td>
</tr>
<tr>
<td>A levels</td>
<td>184 (22.1)</td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td>108 (13.0)</td>
</tr>
<tr>
<td>Master’s degree</td>
<td>147 (17.6)</td>
</tr>
<tr>
<td>Doctorate</td>
<td>14 (1.7)</td>
</tr>
</tbody>
</table>

*RQ1: How Do Users and Nonusers Perceive Opinions of Relevant Groups and Their Influence?*

To get an impression about the distribution of users and nonusers and investigate RQ1, we analyzed the distribution of the participants’ peer groups’ opinions and their perceived influence on the participants (Figures 3 and 4).

*Figure 5* illustrates that the participants’ perception of their peer groups is in general positive, with a higher opinion from media, family doctors, politicians, and virologists/RKI. The perception of social media posts was slightly negative for both users and nonusers. Interestingly, the reported opinions from users for friends and family were way higher than the ones from nonusers; besides, the ones from nonusers were slightly negative. A similar picture was perceived when considering the influence of friends and family.
Figure 3. Distribution of the answers regarding the perceived opinion of different groups in participants’ social environments. W: wave.

Figure 4. Distribution of the answers regarding the perceived influence that different groups in participants’ social environments have on using the Corona-Warn-App. W: wave.
RQ2: What Are the Differences Between Users and Nonusers?

As discussed in the previous section, CWA users seem to perceive their peer groups’ opinions more positively. Thus, we now took up RQ2 and systematically investigated the differences between users and nonusers. The visual impression from Figure 5 is supported by Mann-Whitney tests showing significant differences between users and nonusers except for the opinion of social media postings. According to the means, nonusers generally had a lower mean at both waves (Table 3).

Furthermore, we investigated the influence of gender with a Mann-Whitney U test. The test results indicated that gender does not present any difference in the perception of the peer groups’ opinion when it comes to the opinion of social media posts toward the CWA. The mean was higher for men than for women in both groups (user and nonuser) and in both waves (Multimedia Appendix 2).

We also investigated the influence of age on the perceived opinion and influence of the peer groups. For this purpose, we used a Kruskal-Wallis H test (Table 4). The result showed that the differences were significant between the different age groups for the opinions of virologists/RKI ($P<.03$) and friends/family ($P=.04$), as well as for the influence of the media ($P<.001$), family doctors ($P=.03$), politicians ($P<.001$), virologists/RKI ($P<.001$), and friends/family ($P<.001$). Although there is a tendency within these groups that the oldest group had the highest values, the means do not give a clear picture, as there was another peak for the “40-49-year” age group.
RQ3: How Do the Opinions and the Influence Change Over Time (From Wave 1 to Wave 2) Driven by Infection Rates (Decreased From Wave 1 to Wave 2)?

We also investigated the changes in the perceived opinion and the influence of peer groups over time. Table 5 shows that the differences among users in the first wave and second wave were minimal. However, after applying the Wilcoxon test, we found significant differences:

- Users’ perceived opinion about the media group was significantly lower in the second wave. This holds true for all participants ($P<.001$) and for the users ($P=.001$), but the decrease for nonusers was lower ($P=.15$), and not statistically significant.
- Users’ perceived opinion of politicians was significantly lower for all participants ($P<.001$) as well as for the user ($P=.004$) and nonuser ($P=.003$) subgroups in wave 2.

Table 3. Differences of perceived opinions and influence between users and nonusers

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mann-Whitney significance ($P$ value)</th>
<th>Wave 1</th>
<th>Wave 2</th>
<th>Mann-Whitney significance ($P$ value)</th>
<th>Wave 1</th>
<th>Wave 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opinion of media</td>
<td>&lt;.001</td>
<td>4.71</td>
<td>4.37</td>
<td>&lt;.001</td>
<td>4.83</td>
<td>4.42</td>
</tr>
<tr>
<td>Opinion of family doctor</td>
<td>&lt;.001</td>
<td>5.16</td>
<td>4.40</td>
<td>&lt;.001</td>
<td>5.21</td>
<td>4.39</td>
</tr>
<tr>
<td>Opinion of politicians</td>
<td>&lt;.001</td>
<td>5.34</td>
<td>4.84</td>
<td>&lt;.001</td>
<td>5.46</td>
<td>4.96</td>
</tr>
<tr>
<td>Opinion of virologists/Robert Koch Institute</td>
<td>&lt;.001</td>
<td>5.70</td>
<td>4.98</td>
<td>&lt;.001</td>
<td>5.75</td>
<td>5.04</td>
</tr>
<tr>
<td>Opinion of social media posts</td>
<td>.23</td>
<td>3.83</td>
<td>3.72</td>
<td>.63</td>
<td>3.80</td>
<td>3.68</td>
</tr>
<tr>
<td>Opinion of friends/family</td>
<td>&lt;.001</td>
<td>4.93</td>
<td>3.79</td>
<td>&lt;.001</td>
<td>4.84</td>
<td>3.74</td>
</tr>
<tr>
<td>The influence of media</td>
<td>&lt;.001</td>
<td>4.77</td>
<td>3.98</td>
<td>&lt;.001</td>
<td>4.88</td>
<td>3.97</td>
</tr>
<tr>
<td>The influence of family doctor</td>
<td>&lt;.001</td>
<td>4.59</td>
<td>4.03</td>
<td>&lt;.001</td>
<td>4.59</td>
<td>4.01</td>
</tr>
<tr>
<td>The influence of politicians</td>
<td>&lt;.001</td>
<td>5.00</td>
<td>4.16</td>
<td>&lt;.001</td>
<td>5.09</td>
<td>4.08</td>
</tr>
<tr>
<td>The influence of virologists/Robert Koch Insti</td>
<td>&lt;.001</td>
<td>5.53</td>
<td>4.25</td>
<td>&lt;.001</td>
<td>5.58</td>
<td>4.17</td>
</tr>
<tr>
<td>The influence of social media posts</td>
<td>&lt;.001</td>
<td>4.11</td>
<td>3.77</td>
<td>&lt;.001</td>
<td>4.11</td>
<td>3.75</td>
</tr>
<tr>
<td>The influence of friends/family</td>
<td>&lt;.001</td>
<td>4.81</td>
<td>3.73</td>
<td>&lt;.001</td>
<td>4.81</td>
<td>3.66</td>
</tr>
</tbody>
</table>

Table 4. Opinions and influence with respect to using the Corona-Warn-App for age.

<table>
<thead>
<tr>
<th>Groups/variable based on</th>
<th>Kruskal-Wallis $H$ test in terms of age</th>
<th>Age among users, mean</th>
<th>z</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20-29 years</td>
<td>30-39 years</td>
<td>40-49 years</td>
<td>50-59 years</td>
</tr>
<tr>
<td>Opinion of virologists/Robert Koch Institute</td>
<td>5.55</td>
<td>5.48</td>
<td>5.81</td>
<td>5.79</td>
</tr>
<tr>
<td>Opinion of friends/family</td>
<td>4.78</td>
<td>4.62</td>
<td>4.92</td>
<td>4.79</td>
</tr>
<tr>
<td>The influence of media</td>
<td>4.78</td>
<td>4.67</td>
<td>5.05</td>
<td>4.64</td>
</tr>
<tr>
<td>The influence of family doctor</td>
<td>4.45</td>
<td>4.46</td>
<td>4.65</td>
<td>4.43</td>
</tr>
<tr>
<td>The influence of politicians</td>
<td>5.07</td>
<td>5.01</td>
<td>5.12</td>
<td>4.80</td>
</tr>
<tr>
<td>The influence of virologists/Robert Koch Institute</td>
<td>5.35</td>
<td>5.53</td>
<td>5.56</td>
<td>5.42</td>
</tr>
<tr>
<td>The influence of friends/family</td>
<td>4.59</td>
<td>4.72</td>
<td>4.73</td>
<td>4.69</td>
</tr>
</tbody>
</table>
• Users’ perceived opinion of friends/family with respect to using the CWA had significantly increased for all participants ($P > .001$) as well as for the user ($P = .003$) and nonuser ($P = .02$) subgroups.

• The perceived influence of media toward using the CWA significantly decreased among users ($P = .002$), meaning the influence was weaker but still toward using the CWA.

• The perceived influence of virologists/RKI toward using the CWA significantly increased among nonusers ($P = .01$) toward using the CWA.

• The perceived influence of friends/family toward using the CWA significantly increased among nonusers ($P = .02$) toward using the CWA.

Table 5. Differences in opinion and influence between the 2 points in time (waves 1 and 2).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Wilcoxon significance ($P$ value)</th>
<th>Wave 1, mean</th>
<th>Wave 2, mean</th>
<th>Wilcoxon significance ($P$ value)</th>
<th>Wave 1, mean</th>
<th>Wave 2, mean</th>
<th>Wilcoxon significance ($P$ value)</th>
<th>Wave 1, mean</th>
<th>Wave 2, mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opinion of media</td>
<td>.001</td>
<td>4.83</td>
<td>4.60</td>
<td>.15</td>
<td>4.42</td>
<td>4.32</td>
<td>.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Opinion of family doctor</td>
<td>.18</td>
<td>5.21</td>
<td>5.11</td>
<td>.24</td>
<td>4.39</td>
<td>4.41</td>
<td>.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Opinion of politicians</td>
<td>.004</td>
<td>5.46</td>
<td>5.23</td>
<td>.003</td>
<td>4.96</td>
<td>4.73</td>
<td>&lt;.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Opinion of virologists/RKI</td>
<td>.23</td>
<td>5.75</td>
<td>5.66</td>
<td>.15</td>
<td>5.04</td>
<td>4.92</td>
<td>.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Opinion of social media posts</td>
<td>.18</td>
<td>3.80</td>
<td>3.86</td>
<td>.40</td>
<td>3.68</td>
<td>3.76</td>
<td>.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Opinion of friends/family</td>
<td>.003</td>
<td>4.84</td>
<td>5.01</td>
<td>.02</td>
<td>3.74</td>
<td>3.85</td>
<td>&lt;.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The influence of media</td>
<td>.002</td>
<td>4.88</td>
<td>4.66</td>
<td>.52</td>
<td>3.97</td>
<td>4.00</td>
<td>.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The influence of family doctor</td>
<td>.66</td>
<td>4.59</td>
<td>4.60</td>
<td>.12</td>
<td>4.01</td>
<td>4.05</td>
<td>.53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The influence of politicians</td>
<td>.8</td>
<td>5.09</td>
<td>4.92</td>
<td>.20</td>
<td>4.08</td>
<td>4.23</td>
<td>.90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The influence of virologists/RKI</td>
<td>.08</td>
<td>5.58</td>
<td>5.48</td>
<td>.01</td>
<td>4.17</td>
<td>4.33</td>
<td>.47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The influence of social media posts</td>
<td>.94</td>
<td>4.11</td>
<td>4.11</td>
<td>.59</td>
<td>3.75</td>
<td>3.80</td>
<td>.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The influence of friends/family</td>
<td>.80</td>
<td>4.81</td>
<td>4.80</td>
<td>.02</td>
<td>3.66</td>
<td>3.80</td>
<td>.09</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

RQ4: How Does the Influence of the Relevant Groups Influence the Usage of the CWA?

We used a marginal model with generalized estimating equations to investigate the effect of (non)users’ social environment on the usage of the CWA (Table 6). As can be seen, among social environment variables, the influence of politicians ($P = .04$), virologists/RKI ($P < .001$), and friends/family ($P < .001$) was significant and had changed the usage of the CWA. The other variables were insignificant (media: $P = .13$; family doctor: $P = .80$; social media: $P = .07$; and time: $P = .26$), and their change did not affect the independent variable. The influence of virologists/RKI had the most increasing effect (increasing the odds of using the CWA). Increasing 1 unit of influence of the virologists/RKI variable and keeping the other variables constant increased the odds of using the CWA by 44%. The only decreasing effect (decreasing the odds of using the CWA) was related to the influence of politicians variable. Increasing 1 unit of the influence of politicians variable and keeping the other variables constant decreased the odds of using the CWA by 10%.

We modeled time as a single variable to represent the influence of the different waves to have a simple model and reduce complexity. Modeling it as an interaction term with each of the other independent variables would have resulted in not only a more complex model, but also one with certain overlaps (each variable with its interaction of time), which are hard to interpret.

To investigate how the opinion and the influence of peer groups are correlated, we conducted a Pearson correlation. The Pearson correlation shows that the opinion of the social environment and its influence on using the CWA are related but not strongly correlated (Table 7).
Table 6. Marginal model with generalized estimating equations for the effects of the influence of the social environment on using the app.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>B</th>
<th>Wald chi-square</th>
<th>Significance (P value)</th>
<th>Exp(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-2.641</td>
<td>107.323</td>
<td>&lt;.001</td>
<td>0.071</td>
</tr>
<tr>
<td>Influence of media (print media, websites, film and television)</td>
<td>0.085</td>
<td>2.328</td>
<td>.13</td>
<td>1.088</td>
</tr>
<tr>
<td>Influence of family doctor</td>
<td>-0.012</td>
<td>0.062</td>
<td>.80</td>
<td>0.988</td>
</tr>
<tr>
<td>Influence of politicians</td>
<td>-0.098</td>
<td>4.073</td>
<td>.04</td>
<td>0.907</td>
</tr>
<tr>
<td>Influence of virologists/Robert Koch Institute</td>
<td>0.363</td>
<td>47.024</td>
<td>&lt;.001</td>
<td>1.437</td>
</tr>
<tr>
<td>Influence of social media posts</td>
<td>-0.086</td>
<td>3.261</td>
<td>.07</td>
<td>0.917</td>
</tr>
<tr>
<td>Influence of friends/family</td>
<td>0.308</td>
<td>46.023</td>
<td>&lt;.001</td>
<td>1.361</td>
</tr>
<tr>
<td>Time</td>
<td>0.069</td>
<td>1.291</td>
<td>.26</td>
<td>1.072</td>
</tr>
</tbody>
</table>

**Goodness of fit**

| Quasi likelihood under independence model criterion | NA²  | NA             | NA                    | 1931.927 |
| Corrected quasi likelihood under independence model criterion | NA   | NA             | NA                    | 1938.039 |

²NA: not applicable.

Table 7. The Pearson correlation between opinion toward the social environment and its influence on using the Corona-Warn-App.

<table>
<thead>
<tr>
<th>Correlation</th>
<th>User</th>
<th>Nonuser</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Opinion of media and its influence on using the app</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson correlation (r)</td>
<td>0.397</td>
<td>0.219</td>
</tr>
<tr>
<td>Significance (P value)</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Opinion of family doctor and its influence on using the app</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson correlation (r)</td>
<td>0.395</td>
<td>0.337</td>
</tr>
<tr>
<td>Significance (P value)</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Opinion of politicians and its influence on using the app</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson correlation (r)</td>
<td>0.383</td>
<td>0.156</td>
</tr>
<tr>
<td>Significance (P value)</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Opinion of virologists/Robert Koch Institute and its influence on using the app</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson correlation (r)</td>
<td>0.496</td>
<td>0.252</td>
</tr>
<tr>
<td>Significance (P value)</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Opinion of social media posts and its influence on using the app</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson correlation (r)</td>
<td>0.346</td>
<td>0.324</td>
</tr>
<tr>
<td>Significance (P value)</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Opinion of friends/family and its influence on using the app</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson correlation (r)</td>
<td>0.434</td>
<td>0.377</td>
</tr>
<tr>
<td>Significance (P value)</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

**Discussion**

Following the structure of the previous section, we discuss the RQs one by one.

**RQ1: How Do Users and Nonusers Perceive Opinions of Relevant Groups and Their Influence?**

It is surprising that the perception of participants with respect to their peer groups is in general positive. Given that half of the participants were not using the CWA, we assumed that the perception of the nonusers’ peer groups would be on the lower side of the Likert scale. However, that is only the case for family and friends and social media posts. Social media posts are only perceived to be slightly negative, but the opinion expressed there is perceived lower than from all other groups. This might be related to the ongoing discussions about the Network Enforcement Act (German: Netzwerkdurchsetzungsgesetz, NetzDG), which tries to combat fake news, harassment, and misinformation in social media.
RQ2: What Are the Differences Between Users and Nonusers?
Not surprisingly, users perceive their peer groups’ opinions more positively than nonusers. Gender does not seem to have an influence. However, with increasing age the tendency increased to perceive the peer groups’ opinion more positively. However, there was a peak for the “40-49-year” age group for all peer groups. The “40-49-year” age group overlaps to a large degree with the so-called Gen X, but we could not find any indication that SARS-CoV-2 or technology was perceived differently by this group in Germany.

RQ3: How Do the Opinions and the Influence Change Over Time Driven by Infection Rates?
Hospitalization rates and number of deaths were significantly lower during wave 2 compared with wave 1. The perceived opinion of media and politicians has significantly decreased from wave 1 to wave 2. This fits with the observation that many politicians were (wrongly) blaming the app to be not so useful because it does not send information to the public health departments or blaming data protection for hindering the effectiveness of the app.

The perceived opinion of friends and family as well as their perceived influence toward using the CWA has increased. This might be related to the perception that many public health departments in Germany were overloaded and the official fight against SARS-CoV-2 was given up due to shortage of staff.

RQ4: How Does the Opinion of the Relevant Groups Influence the Usage of the CWA?
The influence of virologists/RKI has the most increasing effect. This could be backed up by a huge presence of the RKI in the media and their decisive role in changing the rules during the pandemic. The only decreasing effect we found was the influence of politicians, which could be explained by the participants getting tired of politicians contradicting each other, feathering their own nest by promoting companies selling masks and other medical equipment to fight the pandemic, and a number of seemingly uncoordinated decisions between the different states and the federation.

Principal Findings
Our results indicate that participants’ perception of their peer groups is in general positive, with a higher opinion from media, family doctors, politicians, and virologists/RKI and a lower, only slightly negative, opinion from social media posts. Users perceived their peer group’s opinion higher than nonusers. A similar pattern can be observed when considering the peer groups’ influence instead of the opinion. The perceived opinion of media and politicians has significantly decreased from wave 1 to wave 2. The perceived opinion of friends and family as well as their perceived influence toward using the CWA has increased. The influence of virologists/RKI has the most increasing effect. The only decreasing effect we found was the influence of politicians.

Limitation
Our study has several limitations. First, our measurement of the opinion and influence of the participants’ peer groups was self-reported. On the one hand, participants might report not only wrong values, but also misinterpret their own perception. On the other hand, this is supported by our evaluation that the participants’ perception is more important than the actual opinion of the peer groups. As a consequence, it is unclear whether lower values stem from a lower perception or a lower opinion (ie, for the report of nonusers).

Furthermore, we can only evaluate correlations but not causality. Therefore, we do not know whether the users’ perception of their peers’ opinion is higher, because they are using the app. In contrast to nonusers, they might be able to identify wrong statements within their peer group and disregard them.

In addition, we only differentiated between users and nonusers. There might be different levels of activity when using the app (ie, participants might just look at infection rates or the personal risk or they could share their own infection).

The separation of groups is not very strict, that is, participants could read statements of the RKI or from other peer groups via social media. However, there is most likely a different perception between those groups; therefore, we had included social media as its own group in the survey.

Our study only had participants located in Germany using the CWA. While the study could not easily be transferred to other countries, as all countries have different contact tracing apps, there might still be cultural influences in the perception of and interaction with the named peer groups. Thus, it could be interesting to have similar investigations in other countries or cultures in the future.

Comparison With Prior Work
To the best of our knowledge, different entities of the social environment of users and nonusers and their influence on the usage of contact tracing apps have not been investigated yet. Only 1 study by Scholl and Sassenberg [52] is related to ours because it investigated the social environment of contract tracing users by measuring a person’s level of identification with the beneficiaries of the contact tracing app (ie, people in their social surroundings) to predict their willingness to use contact tracing apps. Thus, this study is only partially related as it covers only one of the groups we also asked for in our study, namely, friends and family. The authors found that the closer other people in individuals’ social environments are, the more likely they are to use contact tracing apps. This is in line with our finding that a positive opinion and influence of friends and family positively influence the use of the CWA. We contribute to the literature by widening the analysis to different peer groups in the social environment such as doctors or politicians. In addition, our variable social influence is conceptually different from the identification variable in the study by Scholl and Sassenberg [52].

In addition, Oldeweme et al [14] investigated the influence of transparency, social influence, trust in the government, and initial trust in a COVID-19 tracing app on the process of
adopting the app. Their results showed that the transparency dimensions of disclosure and accuracy, in addition to social influence, trust in government, and initial trust, positively influenced the adoption process. They agree on the definition of social influence as the “degree to which an individual perceives that important others believe he or she should use the new system” [5], but they did not investigate which groups were more important than others.

Social influence not only directly influences the adoption of the CWA, but might also influence other important antecedents of the adoption. We have already mentioned that peer groups most likely have an influence on the knowledge of the app [4], which itself influences the privacy concerns, and thus the adoption of the app [3]. Additionally, the perception of the perceived disease threat has been shown to influence the adoption of the app when applying the TAM [70], the Health Belief Model [71], and the Protection Motivation Theory [55]. However, peer groups might also influence the perceived disease threat. Kaspar [55] found that the intention for using a contact tracing app increased when trust in other people’s social distancing behavior decreased. Although other people might be not considered as a peer group, it clearly shows that the perceived behavior of other people influences the adoption of the app. However, Kaspar [55] did not further investigate differences between specific groups’ influence on the adoption. Kostka and Habich-Sobiegalla [13] investigated the public perception toward COVID-19 tracing apps in Germany (and China and the United States) and examined variables such as conspiracy belief (not significant), belief in a second wave (significant), or trust in the state (partially significant). However, they did not investigate the influence of peer groups, although connections have been demonstrated between the COVID-19 pandemic and the 5G conspiracy theory and the spread of misinformation in social networks [72].

Alam et al [73] made use of the Health Belief Model to investigate the public attitude toward vaccinations against COVID-19. They found that, among other factors, “health motivation” was an important factor for the willingness to get vaccinated. Part of this construct is the recommendation of friends, relatives, and the participants’ physician. However, they also did not further investigate the influence of the different groups.

Conclusions and Future Work

Opinions of peer groups play an important role when it comes to the adoption of the CWA. Naturally, not all groups have the same importance. Our results show that the influence of virologists/RKI and family and friends contributed to the adoption of the CWA the most, while politicians only had a slightly negative influence on citizens to use the CWA. Our results indicate that it is crucial to accompany the introduction of such a contact tracing app with an appropriate media campaign with easily understandable technical explanations and the clear approval of political decision makers to support its adoption among a large group of citizens in a given country.

Although the pandemic is considered by many to be overcome, these considerations are still important to make, to create a more resilient society in the future. It is important to investigate not only the adoption of contact tracing apps, but also the adoption of data donation apps. Although the CWA has a feature using which users can report their infections, it would also be beneficial if data could be collected to learn more about the specific disease and how it spreads. For that purpose, not only privacy and privacy concerns should be investigated, but also the influence of peer groups, as they can play a decisive role in the adoption of apps. Besides contact tracing and data donation apps, apps could be used to nudge the users into specific behaviors, such as physical distancing [74], which again would rely on the users’ intention to adopt the app(s).

One natural idea of a future work is to extend our study to other health apps such as those mentioned earlier. One could go even further and investigate health apps such as fitness tracking apps or diet diary apps in general. However, it is also important to scientifically connect the different areas and study the interdependencies of knowledge, perceived disease threat, the opinion and influence of peer groups, and the adoption of the CWA. However, media and misinformation or fake news can influence people’s opinion about the CWA. Therefore, besides a solid education and online/computer literacy, it is important to understand the effects of peer groups to be able to plan and implement governmental information campaigns accordingly.

Acknowledgments

This work was supported by the Goethe-Corona-Fonds from Goethe University Frankfurt and the European Union’s Horizon 2020 research and innovation program under grant agreement number 830929 (CyberSecurity4Europe).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire used in the survey.

[DOCX File, 17 KB - humanfactors_v10i1e45825_app1.docx]

Multimedia Appendix 2
Mann-Whitney U test for gender.
References


69. EUROSTAT 2018. EUROSTAT. 2018. URL: [https://ec.europa.eu/eurostat/de/home] [accessed 2023-05-18]

Abbreviations

CWA: Corona-Warn-App
mHealth: mobile health
RKI: Robert Koch Institute
RQ: research question
TAM: Technology Acceptance Model
Original Paper

Design Validation of a Relational Agent by COVID-19 Patients: Mixed Methods Study

Ashraful Islam¹ ² ³, BSc, MSc, PhD; Beenish Moalla Chaudhry¹, BA, MA, PhD

¹School of Computing and Informatics, University of Louisiana at Lafayette, Lafayette, LA, United States
²Center for Computational and Data Sciences, Independent University, Bangladesh, Dhaka, Bangladesh
³Department of Computer Science and Engineering, Independent University, Bangladesh, Dhaka, Bangladesh

Corresponding Author:
Beenish Moalla Chaudhry, BA, MA, PhD
School of Computing and Informatics
University of Louisiana at Lafayette
James R Oliver Hall
301 East Lewis Street
Lafayette, LA, 70503
United States
Phone: 1 337 482 1676
Email: beenish.chaudhry@louisiana.edu

Abstract

Background: Relational agents (RAs) have shown effectiveness in various health interventions with and without doctors and hospital facilities. In situations such as a pandemic like the COVID-19 pandemic when health care professionals (HCPs) and facilities are unable to cope with increased demands, RAs may play a major role in ameliorating the situation. However, they have not been well explored in this domain.

Objective: This study aimed to design a prototypical RA in collaboration with COVID-19 patients and HCPs and test it with the potential users, for its ability to deliver services during a pandemic.

Methods: The RA was designed and developed in collaboration with people with COVID-19 (n=21) and 2 groups of HCPs (n=19 and n=16, respectively) to aid COVID-19 patients at various stages by performing 4 main tasks: testing guidance, support during self-isolation, handling emergency situations, and promoting postrecovery mental well-being. A design validation survey was conducted with 98 individuals to evaluate the usability of the prototype using the System Usability Scale (SUS), and the participants provided feedback on the design. In addition, the RA’s usefulness and acceptability were rated by the participants using Likert scales.

Results: In the design validation survey, the prototypical RA received an average SUS score of 58.82. Moreover, 90% (88/98) of participants perceived it to be helpful, and 69% (68/98) of participants accepted it as a viable alternative to HCPs. The prototypical RA received favorable feedback from the participants, and they were inclined to accept it as an alternative to HCPs in non-life-threatening scenarios despite the usability rating falling below the acceptable threshold.

Conclusions: Based on participants’ feedback, we recommend further development of the RA with improved automation and emotional support, ability to provide information, tracking, and specific recommendations.

(JMIR Hum Factors 2023;10:e42740) doi:10.2196/42740

KEYWORDS

COVID-19; relational agent; mHealth; design validation; health care; chatbot; digital health intervention; health care professional; heuristic; health promotion; mental well-being; design validation survey; self-isolation

Introduction

Background

A relational agent (RA) is an artificial intelligence (AI)–based computational artifact that is responsible for maintaining a virtual socioemotional relationship with the user for a long time [1]. According to Bickmore and Cassell [2], with an RA, “It’s just like you talk to a friend.” An RA is a special type of conversational agent (CA), which is a conversation system that can process natural language and respond using human language.
CAs are typically implemented as chatbots on the internet or as personal assistants on smartphones or wearable devices. The CA’s interaction with the user does not have to be limited to text. Embodiment of empathy and tangible relational affects transform a CA into an embodied conversational agent (ECA), which is a computer-generated virtual person with animated gestures to facilitate face-to-face interactions between a person and a computer. RAs can also be embodied (embodied RAs) allowing them to use both verbal and nonverbal relational affects over an extended time to form long-term, deep, and meaningful connections with the users. The major difference between RAs and CAs is that, unlike RAs, CAs do not have the capacity to maintain long-term relationships with their users.

Both CAs and RAs have been used for different health care services such as screening, counseling, and caregiving [3-5] in diverse health care settings. Since similar health care services are essential for COVID-19 patients [6-8], CAs and RAs can also be used to support SARS-CoV-2–infected individuals. Moreover, they can remotely deliver essential health services [9], minimizing face-to-face interactions and preventing the transmission of infection. However, in the context of COVID-19, an RA can be more effective than a CA due to its potential to not only offer support and guidance but also maintain a sustained relationship with the user throughout the self-isolation period. A range of CAs was developed during the COVID-19 pandemic, but, to the best of our knowledge, there is no RA that can deliver health care services to COVID-19 patients in non-life-threatening situations.

Objective

The goal of this research was to address this gap by designing an RA-based intervention that can help patients as they go through different stages of COVID-19. In this paper, we present the early usability, usefulness, and acceptance evaluation of a prototypical RA that was designed for COVID-19 patients using a user-centered design (UCD) approach.

Related Work

The majority of virtual agent–based interventions that have been proposed during the COVID-19 pandemic are modeled as CAs, used chatbots, and are capable of performing specific tasks. Some helped patients perform guided symptom checking (eg, [10-12]), while some helped provide mental health interventions [8,13].

Ouerhani et al [14] proposed a cloud-based mobile CA for anxiety-emotion assistance in postquarantine situations during COVID-19 that helped increase consciousness of the real danger of the outbreak. They used natural language understanding (NLU) to analyze and create encouragement among people in infected areas. Also, Welch et al [8] presented an expressive CA that uses automated counseling or motivational interviewing to guide an individual suspected of COVID-19 to reduce stress and inspect thoughts and feelings. Another work by Loveys et al [6] presented a randomized pilot trial of a digital human named “Bella,” a form of virtual assistant (VA) that remotely delivered both stress and loneliness interventions during COVID-19. Completing cognitive behavioral and positive psychology tasks with Bella on a website was part of the intervention. Loneliness, stress, and psychological well-being were all addressed in the activities. Bella was regarded as credible in terms of appearance (human-like facial expression), interpersonal abilities (friendly companionship, nonjudgmental attitude), and information delivery based on the opinions of 30 participants.

Battineni et al [15] proposed an AI-enabled chatbot that can serve patients remotely via awareness and virus updates. It can also provide counseling to help patients recover from psychological damage caused by stress and fear due to the pandemic. Woo et al [16] created “Akira,” an AI-enabled CA, which is close to the work by Battineni et al [15]. It was trained by a deep neural network model, with an accuracy of 90.6% to engage and respond appropriately in 7 forms of pandemic-related conversations, such as mental well-being, cold and flu, medications, and drugs. Akira was tested by 57 people, each of whom came up with a list of 5 questions to ask, and the user experience evaluation revealed that a larger training data set was warranted for better performance.

“Chloe,” a Canadian digital information VA [17], can be portrayed as a benchmark for infodemic management during the COVID-19 pandemic. Chloe would inquire about a user’s symptoms, location, previous travel history, and recent contacts. This evaluation resulted in a tailored suggestion that included connections to local resources such as the COVID-19 recommendations by the user’s province government. Ventoura et al [7] developed an empathy-driven CA named “Theano” that speaks Greek and supports both voice and chat interactions with its users. Theano provides COVID-19 data and facts to users, as well as ideal health practices and the most recent COVID-19–related guidelines. In addition, Theano assists end users in the self-evaluation of their symptoms and guiding them to first-line health care professionals (HCPs). Although Chloe and Theano were responsible for delivering information and guidance, “Jennifer” [18] is another CA that combats misinformation about COVID-19. It answers questions about the COVID-19 pandemic by providing conveniently available and reliable information from reputable sources. It includes a wide range of areas, from case statistics to illness prevention and management best practices.

Methods

Proposed System

We used a UCD approach to develop the proposed RA. Figure 1 illustrates the overall design process.
Intended Users

Overview

The intended user groups for the RA were identified through a literature review, which consisted of a review on COVID-19 prevention guidelines published by the US Centers for Disease Control and Prevention (CDC) [19] and other related manuscripts published in academic journals. An interview with 19 HCPs who had cared for COVID-19 patients during the pandemic was later conducted to confirm initial findings [20]. This led to the identification of 3 user groups (personas) defined in the following sections. An individual named Oli is used to represent each group. We assume Oli has no underlying illness nor medical conditions, and Oli owns a smartphone.

(1) Suspecting Infection

Oli has been experiencing fever, cough, and shortness of breath for many days. Oli is worried that Oli might have caught the SAR-CoV-2 virus but does not want to go for a test unless Oli is sure that it is required. Oli wants to know where to go and what precautions to take in the meantime.

(2) Quarantining at Home

Oli has been diagnosed with COVID-19, but the symptoms are mild. Oli has been advised by the doctors to self-isolate at home for at least 14 days and practice health-promoting behaviors such as consuming a well-balanced diet, getting adequate sleep, and taking prescription drugs. The doctor also tells Oli that Oli can benefit from periodic medical check-ins and vital sign monitoring. Oli also believes that Oli can benefit from someone who can provide support during emergency situations.

(3) Recovering After Infection

Oli has recently recovered from severe COVID-19, which required admission to the hospital and a week in the intensive care unit (ICU). Although Oli is now back at home and recovered from the infection, the memories of the stay at the ICU and a near-death experience are still fresh in Oli’s mind. Witnessing other patients dying in the hospital, the helplessness of the health care workers, and the overall hospital atmosphere have deeply impacted Oli’s psyche. In addition, Oli is experiencing weakness due to the physical damage caused by the virus. Oli feels that people around Oli do not understand Oli’s feelings, and hence, Oli needs emotional support.

User Survey and Task Scenarios

Survey

To identify the intended goals of the target users, we conducted an online survey with individuals (n=21) who had been through the SARS-CoV-2 infection and recovery stages [20]. The participants were asked to enumerate challenges and difficulties they faced during different stages of the disease as well as the support they required to overcome those challenges. The analysis of the survey responses led to the identification of 4 distinct tasks for which the proposed RA needed to be designed (Table 1).

<table>
<thead>
<tr>
<th>User persona</th>
<th>RA’s task scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suspecting infection</td>
<td>• Testing guidance</td>
</tr>
<tr>
<td>Quarantining at home</td>
<td>• Focusing on recovery</td>
</tr>
<tr>
<td></td>
<td>• Handling emergencies</td>
</tr>
<tr>
<td>Recovering after infection</td>
<td>• Postinfection care</td>
</tr>
</tbody>
</table>

Table 1. Identified task scenarios, categorized by each user persona, that the relational agent (RA) should address as a health service delivery intervention.

Figure 1. Sequential stages of overall design process from requirement analysis to prototype development.
(1) Testing Guidance (Scenario 1)

The RA provides testing guidance to Oli when an infection is suspected, by periodically engaging in a dialog to obtain up-to-date symptom status and health metrics.

(2) Wellness Support (Scenario 2A)

The RA provides wellness tips and companionship to Oli during self-isolation at home. The RA also monitors Oli’s symptoms to avert and prevent emergencies.

(3) Handling Emergencies (Scenario 2B)

The RA monitors Oli’s symptoms to avoid and prevent emergencies. Whenever Oli reports an emergency, such as shortness of breath, the RA takes appropriate steps to detect critical situations and connects Oli with the emergency services of a nearby hospital.

(4) Postinfection Care (Scenario 3)

The RA provides companionship and mental health counseling to help Oli recover from the stress of the infection during the recovery phase. The RA attempts to engage Oli in daily activities so that Oli can resume Oli’s pre-infection life.

Since a COVID-19 patient goes through various stages of infection, long-term relationships can help patients successfully navigate this journey. In this work, we present an RA that supports and serves people from the very beginning of their infection (ie, testing guidance) until their postinfection recovery phase (ie, postinfection care). By offering health care advice and required help, the RA functions as a virtual HCP and social companion who attempts to establish a prolonged socioemotional relationship.

Conversation Design

The dialog script for each task scenario was prepared in consultation with HCPs (n=16) who were recruited from the authors’ online social networks. Sample scripts corresponding to relevant task scenarios were first prepared and then sent to the HCPs for feedback via online surveys. Based on thematic analysis, we identified 3 major characteristics required for the dialog scripts: (1) robust and validated screening algorithm, (2) focus on building trust by validating feelings and setting realistic expectations, and (3) long-term relationship building by frequent check-ins, peer support, and psychological interventions. Moreover, the dialogs were influenced by existing diagnosis and coaching models (eg, COVID-19 diagnosis model by CDC [21], posttraumatic stress disorder [PTSD] reduction methods [22]). The details have been reported in a separate publication [23].

Prototype

We developed our prototype using a web-based real-time service called BotSociety [24]. The prototypes developed using BotSociety utilize the XML-based Speech Synthesis Markup Language (SSML) to process natural language and produce responses. BotSociety’s NLU module understands user input obtained via a speech or text recognition module, by comparing it with the information stored in its knowledge base. The computing architecture for BotSociety-based prototypes is summarized in Figure 2. BotSociety prototypes can be distributed for testing and evaluation purposes via a hyperlink, allowing them to be used on any platform connected to the internet via a web browser. This flexibility makes it possible to test user experience on different systems including smartphones, tablet computers, and desktop or laptop computers.

The RA provides verbal and visual (eg, prompts, graphics, animations) information to the user. The user inputs data using voice or touch and clicks. For improved understanding of the communication between the RA and user, all voice interactions are also displayed as text on the RA’s interface. The user can choose to run the prototype on any of the 3 system modes (ie, smartphone, voice assistant, and tablet computer). We used the scripts that were developed in collaboration with the HCPs (domain experts) from the conversation design phase to create the knowledge base of our prototype for each of the task scenarios available in Table 1.

BotSociety provides a flow-based interface for conversation design and a rule-based system that is based on an IF-ELSE dependency mechanism to manage hand-crafted conversations [25] for dialog management. All responses within a category belong to a single topic, and a category is also referred to as a frame. The dependencies allow the RA to shift from one frame to another to ensure the consistency and flow of the conversation. For example, let us assume that the RA wants to coach a user about healthy diet but the user is not interested; the RA will immediately shift the conversation to another frame that is better aligned with user’s current interests. We used the conversation scripts. For example, if a user asks the RA, “Is shortness of breath a symptom of COVID-19?”, the RA will respond “Yes! it is.” and not provide any unnecessary information. The user’s SARS-CoV-2 infection status was obtained via a speech or text recognition module, by comparing it with the information stored in its knowledge base. The RA provides companionship and mental health counseling to help Oli resume Oli’s pre-infection life.
Figure 2. A simplified architecture of the developed prototype in which interactions between the proposed relational agent (RA) and the user take place in human-like natural language. NLU: natural language understanding; SSML: Speech Synthesis Markup Language.

Figure 3. Snippets from the prototype interface is presented in the smartphone view, in which the relational agent (A) checks and shows Oli’s physiological vitals and diagnoses for any COVID-19 symptoms (Scenario-1), (B) helps maintain daily healthy practices for fast recovery while Oli is in home isolation (Scenario-2A), (C) takes action during an emergency condition of Oli and tries to make Oli confident by providing affirmative responses (Scenario-2B), and (D) attempts to engage Oli in daily activities to reduce her posttraumatic stress disorder (Scenario-3).

Prototype Evaluation

Goals

Since we sought to validate the proposed RA design, the goal of this evaluation study was to (1) evaluate the usability of the proposed system, (2) determine people’s perceptions about the RA’s usefulness, (3) determine whether people would accept the RA as an alternative to HCPs in non-life-threatening COVID-19 scenarios, (4) elicit people’s preferences regarding the platforms and devices on which they would want to operate the RA, and (5) improve the system design by obtaining feedback from the participants (design recommendations). The study was targeted toward currently infected and recovered individuals so that all the features or scenarios could be evaluated based on participants’ infection experience and postrecovery constraints. The hyperlink to the developed prototype was included in the survey so that the participants could access it through any web browser available on their smartphones or computers. This evaluation study was conducted online, allowing participants to complete the survey at their convenience. There was no restriction on experiencing the prototype, and the participants were free to access it multiple times. The survey was designed in Google Forms, and the order of tasks in this study was controlled. A hyperlink to the survey was created so that the participants could take part in the study by that hyperlink.
**Participant Recruitment**

To ensure diversity of participants in our study, we posted the survey link on various social media platforms (such as Facebook) and mailing lists. The recruitment had no geographical constraints; therefore, individuals from all over the world could take part in this study. The inclusion criteria were (1) being at least 18 years old, (2) previous or current infection with SARS-CoV-2, (3) a basic understanding of the English language since the prototype was developed in English, and (4) having a minimum knowledge of operating a computer to explore the prototype. Participation was voluntary; there was no restriction on who participated, and no identifying information such as name, email address, nationality, or location was collected. The survey took 45 minutes to 60 minutes to complete, and no incentives were offered for completing the survey. All the participants interacted with the RA through all the available modalities (voice, text, click and touch, prompts).

Our recruitment resulted in 98 (53 men) responses from individuals between 18 years and 64 years old, with a mean of age of 34.42 (SD 11.46) years (n=45, 18-30 years; n=25, 31-40 years; n=17, 41-50 years; n=11, 51-64 years). At the time of the survey, 52 participants were infected (n=25, severe symptoms; n=27, mild symptoms), and 46 participants had recovered from the infection (n=17, severe symptoms; n=29, mild symptoms) on the basis of participants’ self-declarations. Everyone owned a smartphone; 64 participants were Android users, and the remaining were iPhone users. Everyone held at least a high school diploma.

**Study Design**

We used a survey-based evaluation to elicit both quantitative and qualitative feedback about the RA from the participants. The survey began with a brief description of the study and informed consent. Participants were then asked to provide necessary demographics (eg, gender, age range, education level, occupation), infection status (currently infected or recovered), and symptom severity. After completing the demographics section, participants completed 4 study tasks, corresponding to each interaction task scenario (Table 1). Each study task involved, first, reading the patient persona description and then interacting with the developed prototype using provided prompts. In other words, participants were asked to pretend that they were using the RA as the presented patient persona to get help. The prompts and link to the prototype were provided within the survey. After completing the task, participants returned to the survey to respond to a series of questions. Specifically, participants indicated the usefulness of the RA for each scenario on a Likert scale and made suggestions (qualitative) for improving the interaction. Each task took 5 minutes to 10 minutes to complete.

After completing all the tasks, participants completed the System Usability Scale (SUS) [26] and indicated their acceptance of the system on a Likert scale. Participants were also asked to specify on which platforms (eg, smartphones, voice assistants, computers) they would prefer to use the RA.

**Measures and Analysis**

Different types of survey responses were collected, such as Likert scale responses, yes or no responses, and qualitative data. We used the SUS to determine the usability of the presented prototype. The scale was chosen because it is simple to compute and interpret the relative usability and satisfaction with a system. This will also provide a basis for comparison at a later stage of system development.

Microsoft Excel was used to conduct descriptive analysis of the closed questions, which included calculation of means, SDs, percentages, and frequency distributions. The qualitative user feedback was analyzed using the thematic analysis technique [27] using Microsoft Excel. Inductive and deductive coding methods, such as open coding and memoing, were used to code the survey responses after the researchers independently read and reread each response. The researchers then compared their codes and had a discussion to resolve any discrepancies. Related codes were grouped to create minor themes, which were then refined over several iterations and categorized under major themes at the end.

**Ethics Approval**

The institutional review board of the University of Louisiana at Lafayette approved the corresponding user studies (Reference: SP21-82 CACS). All the user studies were conducted online, and the participants responded anonymously. At the beginning of the survey, the study’s objective was stated clearly, and participants could only enter the actual survey after they had indicated a willingness to participate by answering yes to the first question.

**Results**

### System Usability

The mean score of each SUS item is presented in Table 2. The overall average SUS score was 58.82 (SD 10.92). The average SUS score (59.88, SD 11.99) of mildly infected participants was slightly higher than that (57.75, SD 9.89) of severely infected participants. However, the difference between the scores of both groups was not statistically significant according to the Student t test \( P = .06 \).
Table 2. Categorized average System Usability Scale (SUS) scores among participants.

<table>
<thead>
<tr>
<th>SUS item</th>
<th>Mild symptoms, mean</th>
<th>Severe symptoms, mean</th>
<th>Overall sample, mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>I think that I would like to use this system frequently.</td>
<td>4.3</td>
<td>4.05</td>
<td>4.18</td>
</tr>
<tr>
<td>I found the system unnecessarily complex.</td>
<td>3.35</td>
<td>3.05</td>
<td>3.2</td>
</tr>
<tr>
<td>I thought the system was easy to use.</td>
<td>3.95</td>
<td>3.85</td>
<td>3.9</td>
</tr>
<tr>
<td>I think that I would need the support of a technical person to be able to use this system.</td>
<td>3.05</td>
<td>3.5</td>
<td>3.28</td>
</tr>
<tr>
<td>I found the various functions in this system were well-integrated.</td>
<td>3.95</td>
<td>3.65</td>
<td>3.8</td>
</tr>
<tr>
<td>I thought there was too much inconsistency in this system.</td>
<td>3.3</td>
<td>3.2</td>
<td>3.25</td>
</tr>
<tr>
<td>I would imagine that most people would learn to use this system very quickly.</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>I found the system very cumbersome to use.</td>
<td>3.25</td>
<td>3.55</td>
<td>3.4</td>
</tr>
<tr>
<td>I felt very confident using the system.</td>
<td>4.15</td>
<td>4.05</td>
<td>4.1</td>
</tr>
<tr>
<td>I needed to learn a lot of things before I could get going with this system.</td>
<td>3.45</td>
<td>3.2</td>
<td>3.33</td>
</tr>
<tr>
<td>SUS score</td>
<td>59.88 (11.99)</td>
<td>57.75 (9.89)</td>
<td>58.82 (10.90)</td>
</tr>
</tbody>
</table>

*aMean (SD).

Model’s Usefulness

The usefulness of the RA was determined based on participants’ responses to the following question for each scenario and the overall system: “Indicate your degree of agreement with the following: The services provided by the RA are useful for the target persona.” Participants’ ratings of the usefulness of each RA task are presented in Figure 4. For each task (scenario), most of the participants agreed that the system is very useful or useful. Scenario 3 received the smallest number of usefulness votes by both mildly and severely infected participants. There was no severely infected participant who disagreed with the usefulness of any one of the RA tasks. For the overall system, 90% (88/98; SD 7%); severe symptoms: 40/42, 96%; mild symptoms: 48/56, 86%) of the participants thought that the RA is useful. However, among the participants, 9% (9/98; SD 3%; severe symptoms: 2/42, 5%; mild symptoms: 4/56, 8%) were unsure about the model’s usefulness, and 2% (2/98; SD 2%; only mild symptoms: 2/56, 3%) did not think the model was useful.

Figure 4. Chart illustrating the details of participants’ votes (on a 3-point Likert scale) on model usefulness categorized by infection severity: (A) mild symptoms and (B) severe symptoms. Scenario-1: testing guidance; Scenario-2A: wellness support; Scenario-2B: handling emergencies; Scenario-3: postinfection care.

Acceptable Alternative

The acceptance of the proposed RA was measured by participants’ responses to the following question: “Would you accept the proposed RA as an alternative to caregivers/HCPs in non-life-threatening situations?”

Overall, 69% (68/98; SD 6%) of participants (mild symptoms: 40/56, 72%; severe symptoms: 27/42, 64%) agreed that the proposed RA could be an alternative to caregivers or HCPs for non-life-threatening situations. Of the participants with mild symptoms or severe symptoms, 0% (0/56) and 9% (4/42), respectively, declined to accept it as an alternative, and the rest (27/98, 28%) were unsure (ie, neutral).

Preferred Platform

We also asked participants to indicate which platforms (end devices; eg, smartphones, smartwatches, voice assistants [eg, Amazon Alexa, Google Home Mini], and computers [laptops and desktops]) they would prefer for the proposed system. Of the participants, 93% (91/98) chose smartphones, while 17% (16/98) chose smartwatches. Voice assistants and personal computers were ranked in the second and third positions, with 33% (32/98) and 27% (26/98) of the votes, respectively. These
preferences inspired us to develop the target RA as a platform-independent (eg, web-based or cloud-based) application so that the target users can have access to the RA on any of their devices.

**Design Feedback**

**Overview**

We identified 6 significant aspects as a result of the thematic analysis. Briefly, the mildly infected participants wanted to receive more information about the disease and use an interactive interface, whereas the severely infected participants requested tracking and automation within the system. Both groups thought the system could improve in terms of providing emotional support and specific recommendations for managing the disease. The summary (including a few quotes from participants) of identified aspects are provided in the following sections.

(1) **Automation**

Participants thought that the system could be automated to perform additional functions. Particularly, participants wanted the system to automatically determine participants’ health status and take subsequent actions. They specifically urged for an automatic and robust way of handling medical emergencies without the user’s input:

> Since the system has access to bodily measurements, why do users have to initiate the procedure? I guess the system can automatically dispatch a notification to the user alerting about their condition. [P7, Scenario-1]

> In some cases, patients can’t respond, and the guide should act independently to contact emergency services. [P22, Scenario-2B]

(2) **Information**

Participants suggested that the RA should provide users with accurate, up-to-date, and location-specific information. In particular, participants recommended that the RA should be able to provide various kinds of information such as testing appointments and costs:

> News Feed (Regularly updated with new Covid-19 Rules including travel restrictions, safety measures, social gathering restrictions etc.). [P1, Scenario-1]

> Testing strategies vary from country to country; in many places, you can’t just turn up at the test center, but need to phone ahead and make an appointment, or speak to an epidemiologist ... [P5, Scenario-1]

(3) **Tracking**

Participants thought that the RA should be able to track users’ actions and data, so future actions can be informed:

> It could track progression of COVID-19 symptoms so it could help with future diagnosis. [P37, Scenario-1 and Scenario-2]

> Tracking of the location for that day if the person went to hospital for the test or not. [P25, Scenario-1]

(4) **Emotional Support**

Some participants thought the RA would be useful in providing emotional support to patients. In particular, participants thought that the RA could help them feel less lonely during and after recovery, as it gave the impression that someone was looking after them:

> Just wanted to say, I like this feature a lot! The worst part for me was how alone I felt during and after recovery (isolated at home). This kind of thing can be a huge help. [P53, Scenario-2A and Scenario-3]

Participants recommended that the ability of the RA to provide emotional support be further improved. Some participants suggested that the RA should allow patients to seek emotional support from their loved ones. Others suggested connectivity with social media groups:

> Communication with loved ones may help in reducing PTSD. This feature might be included. [P21, Scenario-3]

Other participants thought that the system would not be able to replace a human in terms of providing emotional support:

> I think talking to someone on the phone is more helpful than texting into the system. [P34, Scenario-3]

(5) **Personalized Recommendations**

Participants thought that the RA’s ability to provide more personalized recommendations to the user could be improved. Participants thought that the system should be able to anticipate the user’s situation and provide help and resources accordingly:

> Some medical assistance can be added (eg, medicines, case studies, home remedies) according to a person’s symptoms. [P23, Scenario-2A]

> Infected person normally loses the ability to taste food and smell, some suggestions can be provided in case the person is frustrated about this. [P11, Scenario-1 and Scenario-2A]

Participants also thought that the system should go beyond just recommending to also explaining how these recommendations would be helpful to the patient:

> It’s alright to show what to eat, what type of exercise to do, other precautions to take — it can also be explained why these habits will be helping the COVID-19 patient. [P19, Scenario-2A]

(6) **Interactive Interface**

Some recommendations revolved around making the interface more fun and interactive to increase users’ engagement with the system:

> The exercise tips can be animated instead of still images to make them clear to the person. [P42, Scenario-2A]

Participants also suggested that the proposed RA should be updated to include some entertaining features such as fun games and videos, since patients believed that such features can temporarily diverge patients from the stress of the infection:
Can you add some funny video or fun game. I hope this will help a covid 19 positive people to enjoy some little time with joy to think outside of his or her physical condition. [P15, Scenario-2A and Scenario-3]

Discussion

Principal Findings

This article presents the design validation of a prototypical RA, that targets SARS-CoV-2–infected patients during various stages of the disease. The prototypical RA is the end product of several iterations of finalizing the design requirements (ie, intended users and task identification, interaction dialog design). The RA may also be able to provide COVID-19 patients with immediate assistance. Content and services of the RA may be tailored to address immediate COVID-19 health issues, make recommendations for remedies, and then monitor the patient’s condition.

The COVID-19 outbreak led to the development of a substantial number of CAs or VAs, but research reveals that few of them have been validated in terms of their usability or usefulness or the willingness to be used by patients. For example, Ishii et al [13] only showed the design process of a CA-based companion for people in COVID-19 quarantine. In our work, we addressed these gaps by presenting relevant findings about our proposed RA along with the design process. Our proposed RA has the benefit of being relational in creating relationships with the patients from the very beginning of the infection through the recovery stages when compared with state-of-the-art CA development targeting COVID-19. Other CAs of a similar type have primarily been created to focus on just 1 or 2 COVID-19–related problems or to target a particular COVID-19 stage. For instance, Loveys et al [6] proposed a virtual human–based CA that helps individuals by providing remote intervention for stress and loneliness during COVID-19. Another example is from Siedlikowski et al [17], who presented a CA that acted as a self-assessment tool for COVID-19 diagnosis. The key benefit of our proposed RA is that it merged the 4 possible stages of a COVID-19 patient and addressed them as a whole to deliver health care services depending on the patient’s present and actual situation.

According to a recent user study [28] on designing CAs to overcome COVID-19 difficulties, study participants asked that CAs play a number of important tasks. Participants in that study expressed a desire for CAs to be able to serve as an information hub for guidance regarding COVID-19 diagnosis, a personal assistant for providing health recommendations based on an understanding of patients’ circumstances, and a mental health tool for relieving patients’ stress and feelings of loneliness as they go through the self-quarantine and recovery phases. Participants in the design process also desired that the user personas for the CAs combine comfort and trust, the CAs communicate directly with the patients, and the impact of previous interactions with regard to the infection should be considered when adjusting future interactions with the CAs. The design considerations and needs that the study in [28] reported are all adequately handled in our work, and our proposed RA is capable of fulfilling the participants’ intended demands.

Despite the fact that this was a small-scale validation study, the results offer important valuable insights on how an RA could be designed more efficiently to aid COVID-19 patients. Overall, the majority of participants thought the proposed RA was helpful during the COVID-19 infection and recovery phases. The strength of this study is that the perceptions of the target users (ie, COVID-19 patients) on the proposed RA's design were evaluated at an early development stage.

We view this research as an anchor project to showcase how an RA can handle different scenarios (ie, testing guidance [Scenario-1], encouraging healthy habits at home [Scenario-2A], handling emergencies at home [Scenario-2B], and postinfection care [Scenario-3]) for a similar pandemic. The evaluation suggests that the proposed RA is a promising intervention to address the explored scenarios. Furthermore, patients were willing to accept it as a reliable alternative to HCPs in non-life-threatening situations.

We used the SUS to determine the usability of the presented prototype. The SUS is graded on a scale of 1 to 100, with a higher score indicating higher perceived usability, and the acceptable average value is 68. When compared with similar kinds of CAs [29-31] available in the literature, our proposed RA earned a lower average SUS score; however, the reasons for their higher SUS scores are a longer interaction period (in some cases, more than 20 days) with CAs and the use of the Wizard-of-Oz approach. In addition, providing directions including suggested tasks and talks during user studies supervised their interactions between the CAs and the participants. Familiarity with the prototypes for a longer time and known interactions influenced and contributed to improved perceptions of those systems, which resulted in higher average SUS scores. However, there is greater familiarity with typical CAs such as chatbots and voice assistants than with RAs having embodiment. This factor also led to a lower perceived usability score regarding our proposed RA.

Despite receiving a usability rating below the acceptable score, the prototype received positive feedback from the participants, and they were willing to accept it as an alternative to HCPs in non-life-threatening situations. Both mildly and severely infected participants expressed interest in using the proposed system, and they thought that it would be easy to use and learn. Furthermore, regardless of infection severity, participants were positive about the RA's ability to assist COVID-19 patients. It is noteworthy that all severely infected participants thought that the RA was useful in the present form or could be improved to be useful. No severely infected participant thought that the RA was not useful. The recommendations of each group for improving the interface differed from each other, suggesting that the needs of each group were also different. This makes sense because mildly infected participants were recovering at home, and they were less reliant on medical devices such as a ventilator for recovery. This suggests that we need to pay more attention to understanding the specific needs of mildly and severely infected patients to increase the RA’s appeal.
Participatory design by engaging users during the design and implementation of new technology helps the end product satisfy the needs of its intended stakeholders [32-34]. In line with this fact and the impact of empathy in RA design [35], we engaged the participants to share their views on the final product, and we considered their suggestions and agreed to implement several modifications in the final development of the proposed RA. The ultimate goal is to develop a robust RA-based intervention with maximum usability and efficacy at the time of pandemic situations, particularly by gathering feedback for the RA-enabled system as a social companion in guiding and training the users during various health conditions. These findings are consistent with the prior research explorations described in [3,4,14], and our proposed system performed to a greater extent than the chatbot and ECA-based interventions in [8,15,34-36]. Our findings suggest that, in situations when human intervention is not necessary, patients are willing to receive relevant services from RAs that perform the functions of HCPs, provided the RAs are deliberately and appropriately designed and developed.

Limitations and Conclusion
The study is characterized by an imbalanced ratio between people with mild and severe symptoms. We acknowledge that feedback from more participants with severe symptoms could have provided a more balanced evaluation of the proposed system. Even though participants with severe symptoms participated in the study, their health conditions (eg, weakness, lack of concentration) may have prevented them from responding properly during the study. Finally, because the interactions with the RA were canned and limited, they may not have seemed particularly natural to the participants. Additionally, there is no concrete way to evaluate the interactions between the RA and its users in this study. There are many excellent inventories for assessing patient-doctor or counselor-client interactions [37-39], but there are not enough for evaluating RA-human interaction in health settings. This requirement for a framework to evaluate the dialogs between an RA and its users might be explored by future research.

Only persons who had COVID-19 and recovered were recruited in this evaluation. However, the first scenario (testing guidance) of the RA’s tasks was intended for people who were not sure if they had an infection so they could find out how to test themselves. Since our recruitment led to infected people only, it could be implied that the participants started at the first scenario's stage to confirm they were infected. Based on their past experience with COVID-19 testing, the participants responded to the survey questionnaire for the first scenario. However, a limitation of our study was that none of the individuals who participated had never been infected.

Because of the COVID-19 pandemic, it was not possible to conduct the evaluation study in person. For participants to join remotely and globally while keeping their anonymity, we had to make the survey information and link accessible to the general public. However, now that the pandemic is under control, we plan to do an in-person evaluation study with a new group of individuals by following a standard research protocol.

Participants in the evaluation study did participate anonymously at their discretion, and all responses, such as SARS-CoV-2 infection status, were self-reported. Due to these phenomena, it was possible that the respondents' feedback would not be accurate given that we did not obtain any verification information in order to preserve anonymity.

The work discussed in this paper is the beginning of an intervention that has the potential to serve people during a COVID-19–like pandemic. We are currently developing a high-fidelity version of this tool for in situ evaluation with the target population by addressing the limitations of this work.

Acknowledgments
We thank all the patients and health care professionals who participated in our studies and provided feedback to validate and improve the proposed relational agent. This research did not receive funding support from any external or internal source.

Conflicts of Interest
None declared.

References


Abbreviations

AI: artificial intelligence  
CA: conversational agent  
CDC: Centers for Disease Control and Prevention  
ECA: embodied conversational agent  
HCP: health care professional  
ICU: intensive care unit  
NLU: natural language understanding  
PTSD: posttraumatic stress disorder  
RA: relational agent  
SSML: Speech Synthesis Markup Language  
SUS: System Usability Scale  
UCD: user-centered design  
VA: virtual assistant

Edited by A Kushniruk; submitted 15.09.22; peer-reviewed by MF Kabir, L Guo; comments to author 06.10.22; revised version received 09.10.22; accepted 10.05.23; published 08.06.23.

Please cite as:
Islam A, Chaudhry BM  
Design Validation of a Relational Agent by COVID-19 Patients: Mixed Methods Study  
JMIR Hum Factors 2023;10:e42740  
URL: https://humanfactors.jmir.org/2023/1/e42740  
doi: 10.2196/42740  
PMID:36350760

https://humanfactors.jmir.org/2023/1/e42740  
JMIR Hum Factors 2023 | vol. 10 | e42740 | p.1407

(page number not for citation purposes)
The TeleTriageTeam, Offering Continuity of Personalized Care Through Telemedicine: Development and Evaluation

Janneau Claessens¹, MD; Sigrid Mueller-Schotte²,³, MSc, PhD; Jeannette van Weerden¹, MSc; Helianthe Kort³, MSc, PhD; Saskia Imhof¹, MD, PhD; Robert Wisse¹,⁴, MD, PhD

¹Department of Ophthalmology, University Medical Centre Utrecht, Utrecht, Netherlands
²Department of Optometry and Orthoptics, HU University of Applied Sciences Utrecht, Utrecht, Netherlands
³Research group Technologies for Healthcare Innovations, Research Centre Sustainable and Healthy Living, HU University of Applied Sciences, Utrecht, Netherlands
⁴Xpert Clinics Oogzorg, Zeist, Netherlands

Corresponding Author:
Janneau Claessens, MD
Department of Ophthalmology
University Medical Centre Utrecht
Huispostnummer E03.136
Postbus 85500
Utrecht, 3508 GA
Netherlands
Phone: 31 887551683
Email: j.l.j.claessens@umcutrecht.nl

Abstract

Background: The COVID-19 pandemic taught us how to rethink care delivery. It catalyzed creative solutions to amplify the potential of personnel and facilities. This paper presents and evaluates a promptly introduced triaging solution that evolved into a tool to tackle the ever-growing waiting lists at an academic ophthalmology department, the TeleTriageTeam (TTT). A team of undergraduate optometry students, tutor optometrists, and ophthalmologists collaborate to maintain continuity of eye care. In this ongoing project, we combine innovative interprofessional task allocation, teaching, and remote care delivery.

Objective: In this paper, we described a novel approach, the TTT; reported its clinical effectiveness and impact on waiting lists; and discussed its transformation to a sustainable method for delivering remote eye care.

Methods: Real-world clinical data of all patients assessed by the TTT between April 16, 2020, and December 31, 2021, are covered in this paper. Business data on waiting lists and patient portal access were collected from the capacity management team and IT department of our hospital. Interim analyses were performed at different time points during the project, and this study presents a synthesis of these analyses.

Results: A total of 3658 cases were assessed by the TTT. For approximately half (1789/3658, 48.91%) of the assessed cases, an alternative to a conventional face-to-face consultation was found. The waiting lists that had built up during the first months of the pandemic diminished and have been stable since the end of 2020, even during periods of imposed lockdown restrictions and reduced capacity. Patient portal access decreased with age, and patients who were invited to perform a remote, web-based eye test at home were on average younger than patients who were not invited.

Conclusions: Our promptly introduced approach to remotely review cases and prioritize urgency has been successful in maintaining continuity of care and education throughout the pandemic and has evolved into a telemedicine service that is of great interest for future purposes, especially in the routine follow-up of patients with chronic diseases. TTT appears to be a potentially preferred practice in other clinics and medical specialties. The paradox is that judicious clinical decision-making based on remotely collected data is possible, only if we as caregivers are willing to change our routines and cognitions regarding face-to-face care delivery.

(JMIR Hum Factors 2023;10:e46145) doi:10.2196/46145

KEYWORDS
triaging; telehealth; telemedicine; remote care; ophthalmology; eye care; mobile phone; COVID-19
Introduction

Background

The importance of high-quality remote care was emphasized when most elective hospital care was on hold during the COVID-19 pandemic. The number of regular face-to-face patient consultations were reduced to comply with government-imposed mobility restrictions. Initially, face-to-face in-hospital consultations were considered only when medically urgent. The capacity of our academic outpatient clinic reduced by 90% (from 300 to 30 visitors per day). Before the pandemic, the capacity of ophthalmic care in the Netherlands was already barely sufficient, with accessibility under pressure and ever-growing waiting lists [1]. Future projections offer little perspective, with an estimated increase in national health expenditures from 12.7% of gross national income in 2015 to 19.6% by 2060, owing to our aging society [2]. To address these immediate and future challenges, we conceptualized and executed a novel telemedicine approach, the TeleTriageTeam (TTT).

The TTT is an ongoing collaboration between the HU University of Applied Sciences Utrecht (HU-UAS) and the University Medical Center Utrecht (UMCU). In this approach, a team of undergraduate (ie, bachelor’s degree) optometry students, tutor optometrists, and ophthalmologists worked together to remotely provide eye care [3]. Although originally conceptualized for telephonic triaging and rescheduling appointments during the acute pandemic-related capacity crisis, the approach has evolved into a telemedicine service that included advising patients, refining treatment, or referring patients to other physicians. It appeared highly valuable beyond the acute crisis and is therefore still ongoing. In addition to allowing the continuation of care, the approach created a unique opportunity to continue the training of optometry students during the pandemic while respecting social distancing and quarantining.

Objective

In this paper, we described a novel method of delivering remote care safely and effectively using an innovative approach to interprofessional task allocation and the application of technology for remote vision testing. We aimed to report on the clinical effectiveness of the TTT approach and its impact on waiting lists and discuss its transformation to a sustainable method for delivering remote eye care.

Methods

Synopsis

The TTT approach included evaluations of current (ocular) health status using semistructured anamneses by telephone conducted by optometry students. If visual acuity was of interest for clinical decision-making, patients were requested to perform a remote, web-based eye test in their home environment. Patients were called back after their cases had been discussed by the supervising ophthalmologists, who were responsible for the clinical decision-making.

Process Overview

Eye Care Delivery Before the Pandemic

The UMCU is a tertiary clinic and training institution. Most of the patients in the ophthalmology department have complex and multifactorial eye disorders. New cases typically present after referral by ophthalmologists from regional clinics. After diagnosis and treatment, most of the patients will be referred back to the referring ophthalmologist or the general practitioner (GP) once the condition is stable. Exempts from this policy are complex cases in need of indefinite academic care. Teleconsultations that replaced in-office visits were fairly uncommon, and video consultations were not performed.

Eye Care Delivery During the Pandemic

When the COVID-19 pandemic began in March 2020, about 90% of our outpatient capacity had to be reduced, greatly impacting scheduled appointments and waiting lists. Teleconsultations (ie, telephonic or video-assisted consultations) were preferred to face-to-face in-hospital consultations to limit the number of hospital visits. Patients were referred to the hospital only when medically urgent (eg, neovascular age-related macular degeneration, poorly regulated glaucoma, and retinal detachments). To help prioritize scheduled appointments and restructure the waiting lists in our ophthalmologic department, the TTT was conceptualized. This approach was continued throughout the pandemic, during the various stages of lockdowns and subsequent changes in social restrictions. Shortly after its introduction, TeleTriage became a part of the standard curriculum for the optometry training at the HU-UAS. After a 2-day training program that focused on navigating through the electronic health record (EHR), best practices in data handling, and patient communication, students were enrolled in the TTT program for 4 weeks.

TTT Workflow

Students of the TTT were assigned patients who were on the waiting list or scheduled for ophthalmology resident clinics. First, the students thoroughly studied and summarized the number of hospital visitors. Patients were referred out to the patients by telephone for semistructured anamnesis. Subsequently, they reached out to the patients by telephone for semistructured anamnesis. A triaging checklist was used to assess the current eye health status and identify any changes in general health or medication use. The primary learning task for the optometry students was to make a triaging proposal based on the gathered information, adhering strictly to the existing clinical protocols. The students were supervised by a qualified tutor optometrist and an ophthalmologist. Under Dutch law, the ophthalmologist is responsible and accountable for the final clinical decision. Triaging decisions for the clinical appointment included the following options: maintain, expedite, postpone, cancel, change into a telephone or video consultation with their physician, or refer regionally. Case summaries were recorded in the EHR, and clinical decisions were relayed back to the patients and, if applicable, to the patients’ GPs or the referring ophthalmologists. Figure 1 depicts the workflow of the TTT.
Figure 1. Workflow of the TeleTriageTeam. Optometry students reach out to patients by telephone and make a triaging proposal. A supervising ophthalmologist will make the final clinical decision. The patient will be informed by the students and the decision will be recorded in the electronic health record (EHR). A tutor optometrist will be on site for overseeing the process, assigning patients to the students, and prediscussing proposal options based on current guidelines.

TeleTriageTeam workflow

1. Optometry students
   Case study
   Reach out by telephone
   Remote eye test
   Recording in EHR

2. Ophthalmologist supervisor
   Clinical decision making

3. Optometry students
   Oral or written relay to patients
   Final recording in EHR

Instructor from Utrecht University of Applied Sciences
Prepare and evaluation

The Remote Eye Test

In some cases, patients were requested to perform a remote eye test. This web-based Conformité Européenne (CE)–certified application enables individuals to self-assess their visual acuity in their home environment using their own electronic devices. This test was developed by Easee BV in collaboration with the UMCU and extensively studied in various patient populations [4-7]. To perform it at home, patients need an internet connection, a smartphone, and a computer or tablet. After entering the test via a website on their computer or tablet, users will be instructed to connect their smartphone by scanning a QR code or entering a code sent by an SMS text message. The patients are instructed to stand or sit 3 m from their screen and cover one eye with their hand while the computer or tablet screen displays a sequence of optotypes that the patient should correctly identify. A calibration step ensures that the displayed symbols are correctly sized, regardless of the screen dimensions of the user’s own devices. The smartphone is used as a remote control to submit the answers. At the end of the test, a visual acuity score will be presented (in Snellen decimal system, the common notation to express this outcome in our clinic).

The remote eye test was available via our clinic’s patient portal website. All patients of our clinic have direct access to their medical records via this secured web-based portal. Access is granted through a government-backed identification system (“DigiD”) [8], which ensures data safety and privacy of this digital environment. Patients were directed to the portal to open the eye test via a web link and instructed to write down or save their eye test results after completion. Within the portal, a dedicated questionnaire allowed patients to report their outcomes, after which it became available to the health professionals in the EHR. This manual step was required because the data were not automatically transferred between the remote eye test and the EHR.

Study Population

This study database included all patients who were assessed by the students as part of the TTT project between April 16, 2020, and December 31, 2021. In principle, all patients on waiting lists or with scheduled appointments at the general resident outpatient clinic of the UMCU were screened for eligibility for teletriaging. Patients scheduled for a subspecialty appointment (eg, patients with uveitis and patients referred to pediatric ophthalmology or vitreoretinal consultants) were excluded from the project because of the anticipated complexity of the cases. The consultant ophthalmologists were responsible for downscaling their waiting lists, and these cases are not covered in this paper. Ophthalmologists specializing in corneal pathology were involved in supervising the TTT (depicted in Figure 1 as “ophthalmologist supervisor”); hence, a minority of cases were considered subspecialty cases from the corneal clinic. No further exclusion criteria were applied.

Data Collection

We used real-world clinical data and demographics of the TTT project, gathered by the optometry students, registered in Microsoft Excel (version 16.0.4266.1001 for Windows; Microsoft Corp) and the EHR, HiX (version 6.1; Chipsoft). The characteristics included in the database were as follows: age, sex, diagnosis, date of triaging contact, reachability by phone...
(yes or no), possibility of a video consultation (yes or no), remote eye test offered (yes or no), remote eye test performed (yes or no), triage proposal by the student, and final decision by the ophthalmologist. Free-text variables were recoded into categories before the analysis. Business data on waiting lists and patient portal access were collected from the capacity management team and IT department of our hospital.

**Data Analysis**

The outcomes of this study included the clinical characteristics of the assessed patients, triaging decisions, uptake of the remote eye test, and the effects of triaging on the waiting lists and case mix of our outpatient department. The TTT project had an iterative development to optimize the service. Therefore, interim analyses were performed at different time points during the project, as part of the scheduled project evaluations. This paper presents the synthesis of these analyses.

Statistical analyses were performed using the SPSS Statistics (version 25; IBM Corp). Demographic data, clinical characteristics, and triaging outcomes were available for all included patients (April 16, 2020, to December 31, 2021). These data are descriptively presented as frequencies and percentages and as means and SDs.

Data on patient portal access and uptake of the remote test were available for all patients included up to May 7, 2021. The data are descriptively presented as frequencies and percentages. The differences in age between active and nonactive patient portal users, invited and uninvited for the remote eye test, and successful and unsuccessful performance of the remote eye test were compared using the independent samples 2-tailed $t$ test. Age differences between the groups were considered statistically significant at a $P < .05$.

**Ethical Considerations**

An anonymized, coded version of the TTT database was used to analyze the clinical data, precluding the research team from tracking patients on an individual level. Analyses were performed in accordance with Dutch privacy laws and the Declaration of Helsinki in the context of quality control and health care evaluation. According to national regulations (Centrale Commissie Mensgebonden Onderzoek), ethics approval and informed consent are not required when the quality of a novel health care delivery system is investigated for local applications [9].

**Results**

**Population Characteristics**

Our database included 3658 registrations of cases that were assessed in this project. The clinical characteristics and demographics of the assessed patients are summarized in Table 1, and this distribution reflects the general outpatient clinic population of our academic hospital. Sex distribution among the patients was equal (female patients: 1902/3658, 52%). The mean patient age was 59 (SD 19) years. The most frequent diagnosis categories were “corneal and conjunctival diseases” (632/2527, 25.01%), “glaucoma” (432/2527, 17.1%), and “screening for ophthalmic disease” (322/2527, 12.74%). The latter included routine screening of patients who had systemic diseases and used chronic medication with an increased risk of ocular disease (eg, protocolled hydroxychloroquine screening).
Table 1. Demographics and clinical characteristics of patients assessed by the TeleTriageTeam.

<table>
<thead>
<tr>
<th>Values</th>
<th>All cases (N=3658)(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1756 (48)</td>
</tr>
<tr>
<td>Female</td>
<td>1902 (52)</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
</tr>
<tr>
<td>Value, mean (SD)</td>
<td>59 (19)</td>
</tr>
<tr>
<td>Value, range</td>
<td>11-97</td>
</tr>
<tr>
<td><strong>First-year cohort (n=2527)(^b)</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Diagnosis category(^c), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Corneal and conjunctival diseases</td>
<td>632 (25.01)</td>
</tr>
<tr>
<td>Glaucoma</td>
<td>432 (17.1)</td>
</tr>
<tr>
<td>Screening for ophthalmic disease</td>
<td>322 (12.74)</td>
</tr>
<tr>
<td>Screening for diabetic retinopathy</td>
<td>269 (10.65)</td>
</tr>
<tr>
<td>Cataract and other lens abnormalities</td>
<td>266 (10.53)</td>
</tr>
<tr>
<td>Retinal and macular diseases</td>
<td>225 (8.9)</td>
</tr>
<tr>
<td>Eye lid and orbit pathologies</td>
<td>120 (4.75)</td>
</tr>
<tr>
<td>Neuro-ophthalmological diseases</td>
<td>110 (4.45)</td>
</tr>
<tr>
<td>Other (eg, refractive errors or unspecified vision loss)</td>
<td>97 (3.84)</td>
</tr>
<tr>
<td>Uveitis</td>
<td>88 (3.48)</td>
</tr>
<tr>
<td>Pathologies of the bulbus, sclera or vitreous</td>
<td>75 (2.97)</td>
</tr>
</tbody>
</table>

\(^a\) All consecutive cases assessed between April 16, 2020, and December 31, 2021.
\(^b\) All consecutive cases assessed between April 16, 2020, and April 7, 2021.
\(^c\) Diagnosis categories are based on “Diagnosis Treatment Combinations”. The Diagnosis Treatment Combinations coding is the Dutch registration method for charging health care to the insurer or the patient.

**Triaging Outcomes**

The triage outcomes are presented in Table 2. For approximately half (n=1789, 48.91%) of the 3658 assessed cases, an alternative to the conventional face-to-face consultation was found. The appointment was cancelled in 212 (5.8%) of the cases, or postponed 733 (20.04%) times, with a mean delay of 6 (SD 4) months. Of the total 3658 patients, the face-to-face consultations of 132 (3.61%) patients were changed to teleconsultations with the ophthalmologist. A substantial proportion of patients (492/3658, 13.45%) was dismissed from academic care, as there was no solid ground for specialized follow-up. Other decisions included consulting with other specialists (194/3658, 5.3%).

Table 2. Triaging outcomes based on the final clinical decision made by ophthalmologists.

<table>
<thead>
<tr>
<th>Triaging outcomes</th>
<th>All cases (N=3658)(^a), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consultation unchanged</td>
<td>1869 (51.09)</td>
</tr>
<tr>
<td>Consultation postponed(^b)</td>
<td>733 (20.04)</td>
</tr>
<tr>
<td>Consultation expedited</td>
<td>26 (0.71)</td>
</tr>
<tr>
<td>Consultation cancelled</td>
<td>212 (5.8)</td>
</tr>
<tr>
<td>Consultation changed to teleconsultation</td>
<td>132 (3.61)</td>
</tr>
<tr>
<td>Referral to regional hospital or general practitioner</td>
<td>492 (13.45)</td>
</tr>
<tr>
<td>Other</td>
<td>194 (5.3)</td>
</tr>
</tbody>
</table>

\(^a\) All consecutive cases assessed between April 16, 2020, and December 31, 2021.
\(^b\) Mean delay 6 (SD 4) months.
The other half (1869/3658, 51.09%) of the patients still required the scheduled face-to-face examination at the clinic and were marked as “maintain the consultation.” The consultations of a few patients (26/3658, 0.71%) were expedited after noting warning signs in the telephonic anamnesis.

### Access to the Patient Portal and Uptake of the Remote Eye Test

Interim analyses of patient portal access and remote eye test performance were conducted 1 year after the start of the project in May 2021. Most of the assessed patients (1667/2634, 63.3%) up to this date were “active” users of the patient portal, meaning they had logged on to this web service at least once. These active users were, on average, slightly younger than those who did not access (mean age 55, SD 18 years vs mean age 65, SD 18 years, respectively; \( P < .001 \)). Patient portal use decreased with age; 75.7% (390/515) of the patients who were aged <40 years were active users, whereas only 32.1% (97/302) of the patients who were aged ≥80 years used the service, as shown in Table 3.

#### Table 3. Access to the patient portal and remote eye test, stratified per age category.\(^a\)

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>Active users of the patient portal, n (% of category total)</th>
<th>Invited to perform the remote eye examination, n (% of active users)</th>
<th>Successful completion of the remote eye test, n (% of invited patients)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All categories (n=2634)</td>
<td>40-49 (n=234, 8.9%)</td>
<td>50-59 (n=409, 15.5%)</td>
</tr>
<tr>
<td>≥80 (n=302, 11.5%)</td>
<td>1667 (63.3)</td>
<td>390 (75.7)</td>
<td>170 (72.6)</td>
</tr>
<tr>
<td>70-79 (n=591, 22.4%)</td>
<td>301 (73.6)</td>
<td>395 (67.8)</td>
<td>314 (53.1)</td>
</tr>
<tr>
<td>60-69 (n=583, 22.1%)</td>
<td>314 (53.1)</td>
<td>184 (11)</td>
<td>97 (32.1)</td>
</tr>
<tr>
<td>50-59 (n=409, 15.5%)</td>
<td>17 (10)</td>
<td>43 (11)</td>
<td>19 (44.2)</td>
</tr>
<tr>
<td>40-49 (n=234, 8.9%)</td>
<td>38 (12.6)</td>
<td>17 (10)</td>
<td>7 (41.2)</td>
</tr>
<tr>
<td>&lt;40 (n=515, 19.6%)</td>
<td>301 (73.6)</td>
<td>19 (50)</td>
<td>19 (44.2)</td>
</tr>
<tr>
<td>All categories (n=2634)</td>
<td></td>
<td></td>
<td>23 (44.2)</td>
</tr>
<tr>
<td></td>
<td>184 (11)</td>
<td></td>
<td>14 (48.3)</td>
</tr>
<tr>
<td></td>
<td>82 (44.6)</td>
<td></td>
<td>10 (0)</td>
</tr>
</tbody>
</table>

\(^a\)Cross-sectional analysis based on data from University Medical Center Utrecht IT department in May 2021. The patients who were not actively using the patient portal were, on average, older than the patients who had been using the patient portal (mean age 65, SD 18 years vs mean age 55, SD 18 years; \( P < .001 \)).

Table 3 indicates that 11% (184/1667) of the active patient portal users were invited to perform the eye test. The eye test was offered if three conditions were met: (1) patients had access to the patient’s portal, (2) patients had access to a smartphone and computer or tablet, and (3) visual acuity was of interest in clinical decision-making. The reasons for not inviting patients to perform the test were not registered. Invited patients were only slightly younger than those who were not invited (mean age 53, SD 18 years vs mean age 59, SD 19 years; \( P < .001 \)). Among the invited patients, there was no significant difference in age between the patients who successfully performed the eye test after the invitation and those who did not (\( P = .91 \)).

### Waiting List Reduction at the Outpatient Department

To evaluate the impact of TeleTriage on our clinic’s health care delivery, we performed several cross-sectional investigations. Figure 2 illustrates the number of patients on the waiting lists in the outpatient department of the ophthalmology residents at our clinic and the pandemic-related restrictions, as expressed by the clinic capacity and lockdown stringency (as reported using the Oxford COVID-19 Stringency Index) [10]. The first blue bar effectively represents the pre-pandemic status of the waiting list, as the government’s lockdown measures were only effective from March 15, 2020, and waiting list effects needed some time to accumulate and materialize. During the following months, the waiting lists grew, peaking in summer 2020 (n=1991, July 2020). TeleTriage assisted tremendously in the continuation of the most urgent care because the clinic capacity was markedly reduced during this first lockdown (~80%). The lockdown stringency varied over time. Lockdown restrictions were lessened in summer 2020, coupled with a slight normalization of clinic capacity (approximately ~30%).
The team productivity of the TTT peaked in July 2020, when approximately 500 cases were assessed over the month. Productivity mostly relied on the number of optometry students assigned to the project and the number of tutor optometrists. From September 2020 onward, TeleTriage became part of the standard curriculum of the optometry training at the HU-UAS, leading to a stabilized inflow of student optometrists. Naturally, other factors influenced the waiting list, such as the fluctuating number of new referrals or surgical capacity. Both dropped during the first months of the initial COVID-19 lockdown but normalized during 2020, albeit at a slightly lower level than in 2019 [11]. Since the end of 2020, we have managed to balance the influx of new patients and referrals to our general outpatient clinic with our reduced capacity and the added outflow of patients owing to TeleTriage, even with periods of imposed restrictions (as reflected by the increasing Stringency Index).

Developing TeleTriage Into a Tool for Value-Based Health Care Delivery

During the initial global COVID-19 lockdown, TeleTriage served to retain continuity of care for the most urgent eye care. No patients with an urgent need for eye care were denied service within the TTT, including retinal detachments, progressive glaucoma, and wet age-related macular degeneration. Within months, our clinic’s capacity recuperated, and the TTT allowed us to process the backlog of patients awaiting an appointment (almost 2000 patients at its peak in summer 2020). Patients with lower urgency or complexity were often processed completely remotely and had their face-to-face consultations cancelled or postponed or were referred externally.

Interestingly, TeleTriage offered a potent means to judiciously select patients for nonspecialty follow-up with regional ophthalmologists or GPs. Referrals were customized to a high degree, with personal telephone feedback and a tailored written medical summary provided to both patients and caregivers. As a result, our patient population slowly but steadily grew more academic with less protocolled care of higher complexity.

A business analysis showed that the eye care delivered by our outpatient clinic in 2021 better adhered to the parameters of the academic care set in 2019. First, in 2021, registrations of Diagnosis Treatment Combinations fell significantly more often within our defined academic care profile (+15% increase compared with 2019). Second, the care delivered was significantly more often considered a strategic theme of the department (+14% increase compared with 2019). Our academic care profile is defined at the institutional level as tertiary referred care, pertaining to hospital-wide strategic themes, or last-resort care. Strategic themes are defined at the department level and indicate when certain Diagnosis Treatment Combinations are compliant with the vision of our management team and adhere to the spearpoints of the UMCU. Note that TTT only considered the general, glaucoma, and cornea outpatient clinics (25% of total patient volume) and not the other subspecialty clinics such as surgical, medical retina, uveitis or orbit, neuro-ophthalmology, and pediatrics. These analyses could only be made for our eye clinic as a whole, with an average of 8000 patients on the waiting list. Interestingly, the TTT still exerted a substantial effect on our overall case mix, whereas the addressable population was only 25% of our total clinic. Should one hypothetically apply this effect to all eye care patients, the case mix changes are assumed to be even more pronounced.
Discussion

Principal Findings

We present a novel method to triage eye care patients remotely, using interprofessional collaboration, teleconsultations, and remote vision testing. This project was conceptualized and catalyzed by the sudden COVID-19 pandemic. Subsequently, we developed it as a tool to deliver value-based health care beyond the primary pandemic setting. This innovation was successful in reducing approximately half of the planned care while providing continuity of care for the most urgent cases and deferring or cancelling consultations judiciously or referring the remaining patients after obtaining a specialist’s consideration. The TTT has helped mitigate the backlog of waiting lists that had been built during the initial months of the pandemic. Limited resources were required, and to the best of our knowledge, this telemedicine approach was the first of its kind to actively involve optometry students and remote eye testing in the workflow. Student participation is beneficial for teaching and training, but it can also enable a high turnover without additional staffing. To date, we continue to use this method in our department, as it offers a tool for value-based health care, delivering “the right care in the right place” (“de juiste zorg op de juiste plek” [12]), and is timely when relevant and needed. We propose that similar workflows could be conceptualized in other eye clinics with more GP referrals (eg, regular hospitals and specialized eye clinics), as less-complex pathology appeared easier to triage. However, this could be offset by a lack of available clinical data; in this project, we often had extensive patient charts at our disposal with numerous auxiliary investigations. Other medical specialties could similarly benefit from a working method as described here and contribute to the human capital challenges in both health care delivery and health care education, only if there is availability of technology for remote assessments and delegated personnel to interpret and collect these data [13-15]. Advanced eHealth technologies are not always required. Our project demonstrated that most triaging decisions were based on the clinical information collected by phone in addition to the data already available in the EHR.

When delivering remote care and triaging services, several ethical considerations and challenges should be considered. One major challenge is to deliver care that is noninferior to a face-to-face examination in terms of quality and safety. In this project, all optometry students worked under supervision, and none of their decisions were made independently. Although quality and safety aspects could not be examined in our descriptive analyses, we argue that in any form, clinical decision-making on available data by ophthalmologists is of the highest quality and pace when compared with other eye care professionals. Our asynchronous method, in which patients were called back after the case discussion with the supervising ophthalmologist, allowed ample time to thoroughly review patient health records and look into national and international guidelines and peer-reviewed literature. In addition to improving quality of care, this method also creates excellent teaching opportunities with real exposure to patient communication and clinical decision-making. Moreover, a plenary discussion of the case summaries was helpful for our ophthalmology residents. Incidentally, the case review resulted in the planning of additional examinations before the planned consultation, thereby improving the efficiency.

Importantly, our remote triaging arguably increased the safety of our population when compared with no care at all. Although this may sound as clear as day, the latter is an inconvenient reality for patients who spend weeks or months waiting for their appointments. In utopia, without restrictions on the amount of care we can deliver, we would happily invite everyone to our clinic for a specialist examination. In reality, scarcity of time and means demands alternative solutions to deal with the ever-increasing waiting lists for routine eye care that further soared during the pandemic. Access to eye care is of paramount importance, and TeleTriage is a novel approach to improve the access.

The biggest lesson learned during this project is that clinical decision-making is often possible without seeing patients in the clinic, especially during the routine follow-up of patients with chronic diseases. A judicious decision to cancel or postpone the consultations or refer patients to specialists could be made frequently based on the patients’ history, current complaints, home-assessed visual acuity, and knowledge of disease biology and epidemiology. However, at an almost equal rate, our ophthalmologists concluded that patient safety could be compromised when further delaying care and decided to maintain (or expedite) consultations at the clinic. The most commonly encountered reasons were the nature of the disease (eg, asymptomatic diseases such as progressive glaucoma), red flags in the case summary (eg, poorer visual acuity or vision symptoms), a lack of trust in the case summary (eg, inconsistencies or language barriers), or existing protocols mandating follow-up (eg, screening for hydroxychloroquine maculopathy or diabetic retinopathy). Naturally, important clinical findings such as ophthalmic examinations, intraocular pressure assessments, and optical coherence tomography imaging can only be assessed in person. For only a small fraction (4%) of the cases, the scheduled face-to-face consultation was changed into a telephonic consultation with their ophthalmologist, although it should be noted that treatment advice or feedback was often delivered via the TTT approach itself. These cases are de facto teleconsultations, paving the way for cancellation or postponement of consultations without compromising quality of care. In this way, while originally conceptualized for triaging, the TTT approach evolved into a telemedicine service that included the full assessments of patients and remote health care delivery.

The supposedly decreased human interaction between patients and their health care professionals is another ethical concern that is frequently introduced as an argument against telemedicine implementation. In this project, we experienced little resistance and no formal complaints from patients who were contacted by phone. The extraordinary situation of the pandemic eased the acceptance of this project, though we also consider the personalized approaches and tailored communication between the students and the patients a vital reason for success. A few patients opposed the final clinical decisions. This was most frequently encountered when patients had a long-standing
relationship with our hospital and were referred to another eye care provider. Invariably, not all patients suitable for referral were referred. The data reported in this study reflect the final management rather than the initially proposed management. Differences between these 2 were not recorded; therefore, the true size of this effect could not be quantified.

Technology adoption is another challenge in the delivery of remote care. Not all patients are willing or able to use telemedicine services. In this project, most clinical decisions were based on the information gathered via phone. More or less all of our patients had access to telephone services, so we did not encounter technical difficulties while collecting these data. In addition, a platform for remote eye examinations was available to the team to collect quantifiable information about the visual function of the patients. As this service requires access to technology and basic digital skills, adoption issues arose. The proportion of internet access in the Netherlands has been reported to be the highest in Europe: 98% of households had direct internet access in 2019 [16]. Nevertheless, digital literacy is age associated and related to the technological competencies that were required during the life course [17,18]. Internet use is less common among the older generations [19]. Most of our ophthalmic population was older. Despite the high internet accessibility and—relatively high—digital literacy rate in the Netherlands compared with other European countries, the uptake of this eHealth application and its role in clinical decision-making was rather low for 2 reasons. First, the students did not invite all patients to perform the eye examination. Obviously, a quantifiable visual acuity outcome is not always essential for clinical decision-making, and unfamiliarity of new team members with the web-based platform reduces professional adaptation. More importantly, the lack of patient portal access and initial resistance of some patients to perform the computer-based test were the evident barriers that precluded the students from guiding the patients through the examination. Second, approximately half of the invited patients did not complete the eye examinations. Anecdotal telephonic feedback from patients who did not complete the test was collected by the research team (JC). Patients reported that the clinic’s patient portal environment was difficult to navigate. Frequently, there were problems with manually entering the test outcomes into the questionnaire. To a lesser degree, a lack of time or motivation was reported. Interestingly, the instructions for the eye test itself were reported to be clear. This is in line with a recently published study on cataract patients (mean age 70, SD 7 years) [20]. In-depth interviews and quantitative questionnaires based on Technology Acceptance Models identified an overall positive attitude toward the web-based eye test. We propose that better integration of this test into the patient portal will make it easier for patients to access the tool and, more importantly, will waive the need to manually enter one’s outcomes. Positive experiences are expected to increase staff confidence in inviting patients to perform the examination. Engaging patients in self-measurements can promote self-awareness, self-management, and ownership of one’s health and well-being. This complies with the transition to patient-centered care models, as reported in a World Health Organization report on eHealth implementation [21].

Eye Care Delivery After the Pandemic
Changing demographics, increased technical possibilities, and a higher prevalence of systemic disorders with ocular manifestations (eg, diabetes) are expected to drastically increase the future demand not only for ophthalmic care [22,23] but also for other domains of health care [2]. In the Netherlands, it is estimated that by 2060, one in 3 people should be working in the health care industry to tackle these demands. As this is not feasible, alternative strategies are required to maintain health care accessible for all, such as prioritizing and improving efficiency [2]. Therefore, we propose that the TTT approach is highly valuable beyond the pandemic setting and of great interest for future purposes.

An important aspect of this project was that the practice pattern was preliminarily introduced within a short period. Our approach could be extended by enriching the remote monitoring platform with options for obtaining images remotely. In the United Kingdom, more evolved triaging workflows have been very successful in reducing hospital visits while maintaining communication, patient safety, and clinical quality, even before the pandemic [24-27]. Especially for retinal disorders, diagnosis and treatment rely increasingly on optical coherence tomography imaging devices rather than fundoscopy [28], which allows an asynchronous approach to diagnostics and clinical decision-making. Therefore, several eye clinics have started to refer patients to remote community clinics for obtaining these images. As not all screened retinas require treatment or further examination, this significantly reduces the burden of clinic visits. Interestingly, an added beneficial effect was higher attendance of diabetic retinopathy screening based on a telemedicine-based methodology when compared with conventional screening [29]. The combined approach of remote diagnostics with centralized asynchronous augmented intelligence clinical decision-making certainly holds promise for the future; this TeleTriage project provides important lessons in this regard. We hope that this manuscript inspires (young) colleagues to rethink how eye care is delivered and that it provides insights into how to become architects of this change. Future studies could focus on further exploring patients’ perspectives and cognitions on teletriaging, analyze clinical outcomes and safety aspects, and evaluate the cost-effectiveness of this telemedicine approach.

Conclusions
In conclusion, our novel approach to remotely review cases and prioritize urgency has been successful in maintaining continuity of care despite the COVID-19 pandemic. The project evolved into a telemedicine service of great interest for future purposes, as it aligns with the current trends toward remote care delivery and reduces the burden of hospital visits. The asynchronous triaging allows efficient task allocation without compromising the quality of care, as medical specialists are responsible for the final clinical decisions. The paradox debated in this paper is that judicious clinical decision-making based on remotely collected data actually is possible, only if we as caregivers are willing to change our routines and cognitions regarding face-to-face care delivery. Patient acceptance of this novel method of care delivery is essential for success and is promoted.
by individual communication and tailored clinical decision-making (ie, patient-centered care). In addition, the triaging method has been highly valuable for educating future health care professionals in understanding the course of disease, communicating with patients, and clinical decision-making. This project serves as a proving ground for upcoming innovations in remote eye care delivery and could play a comparable role for other clinical domains.

Acknowledgments

The authors are grateful to the ophthalmology residents of University Medical Center Utrecht who assisted in overseeing the data collection and preparing and analyzing data for interim evaluations. Furthermore, the authors thank all participating optometry students and supervising optometrists from HU University of Applied Sciences Utrecht for collecting the data and their continued feedback on fine-tuning the TeleTriageTeam (TTT) processes.

This research did not receive any specific grants from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability

The data sets generated and analyzed during this study are not publicly available, as the paper is based on real-world clinical data and interim analyses were performed in the context of quality control and quality assurance, but are available from the corresponding author upon reasonable request.

Authors' Contributions

RW, SM-S, JvW, and SI contributed to the development of TeleTriageTeam. RW, SM-S, JC, and HK contributed to the conceptualization of the paper. JC and SM-S contributed to data analyses. JC contributed to manuscript drafting. All authors contributed to critical revision of the manuscript and final approval.

Conflicts of Interest

RW is the chief medical advisor at Easee BV, Amsterdam, the Netherlands. He was not involved in data collection or analysis in this study. All other authors declare no other conflicts of interest.

References

Abbreviations

CE: Conformité Européenne
EHR: electronic health record
GP: general practitioner
HU-UAS: HU University of Applied Sciences Utrecht
Claessens et al

TTT: TeleTriageTeam
UMCU: University Medical Center Utrecht

Please cite as:
The TeleTriageTeam, Offering Continuity of Personalized Care Through Telemedicine: Development and Evaluation
JMIR Hum Factors 2023;10:e46145
URL: https://humanfactors.jmir.org/2023/1/e46145
doi:10.2196/46145
PMID:37311121

©Janneau Claessens, Sigrid Mueller-Schotte, Jeannette van Weerden, Helianthe Kort, Saskia Imhof, Robert Wisse. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 28.07.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
The Effects of a Health Care Chatbot's Complexity and Persona on User Trust, Perceived Usability, and Effectiveness: Mixed Methods Study

Joshua Biro1*, MS; Courtney Linder1*, MS; David Neyens1, MPH, PhD
Department of Industrial Engineering, Clemson University, Clemson, SC, United States
*these authors contributed equally

Corresponding Author:
David Neyens, MPH, PhD
Department of Industrial Engineering
Clemson University
100 Freeman Hall
Clemson, SC, 29634
United States
Phone: 1 8646564719
Email: dneyens@clemson.edu

Abstract

Background: The rising adoption of telehealth provides new opportunities for more effective and equitable health care information mediums. The ability of chatbots to provide a conversational, personal, and comprehensible avenue for learning about health care information make them a promising tool for addressing health care inequity as health care trends continue toward web-based and remote processes. Although chatbots have been studied in the health care domain for their efficacy for smoking cessation, diet recommendation, and other assistive applications, few studies have examined how specific design characteristics influence the effectiveness of chatbots in providing health information.

Objective: Our objective was to investigate the influence of different design considerations on the effectiveness of an educational health care chatbot.

Methods: A 2x3 between-subjects study was performed with 2 independent variables: a chatbot’s complexity of responses (eg, technical or nontechnical language) and the presented qualifications of the chatbot’s persona (eg, doctor, nurse, or nursing student). Regression models were used to evaluate the impact of these variables on 3 outcome measures: effectiveness, usability, and trust. A qualitative transcript review was also done to review how participants engaged with the chatbot.

Results: Analysis of 71 participants found that participants who received technical language responses were significantly more likely to be in the high effectiveness group, which had higher improvements in test scores (odds ratio [OR] 2.73, 95% CI 1.05-7.41; P=.04). Participants with higher health literacy (OR 2.04, 95% CI 1.11-4.00, P=.03) were significantly more likely to trust the chatbot. The participants engaged with the chatbot in a variety of ways, with some taking a conversational approach and others treating the chatbot more like a search engine.

Conclusions: Given their increasing popularity, it is vital that we consider how chatbots are designed and implemented. This study showed that factors such as chatbots’ persona and language complexity are two design considerations that influence the ability of chatbots to successfully provide health care information.

(JMIR Hum Factors 2023;10:e41017) doi:10.2196/41017

KEYWORDS
electronic health record; EHR; health information; health education; patient education; chatbot; virtual agent; virtual assistant; usability; trust; adoption; artificial intelligence; effectiveness

Introduction

As health care technology advances, internet usage increases, and cultural norms shift (eg, in response to the COVID-19 pandemic), people are receiving more health care information from virtual mediums (eg, telehealth) than ever before [1]. This rising adoption of telehealth provides new opportunities for more effective and equitable health care information mediums.
One such promising health care information medium is chatbots. Chatbots provide a conversational, personal, and comprehendible avenue for learning about health care information. The conversational aspect of chatbots has been shown to help support people in online groups for various health conditions [2]. The personal aspect of chatbots has been shown to excel at providing information on sensitive topics, such as sex-, drug-, and alcohol-related questions of young adults, as chatbots are perceived to be faster and more anonymous than conventional search engines for discussing these sensitive issues without judgment [3]. The comprehensible aspect of chatbots is perhaps their greatest asset for health care applications, as chatbots have been shown to be a more effective resource for finding health care information than conventional internet-based searching for individuals with low health literacy [4]. Health literacy is crucial for empowering people to manage their health [5], yet most health information is written at levels that exceed people’s understanding [6]. This disconnect between health literacy and health information is estimated to cost the United States’ health care system between US $106 billion and US $238 billion annually [7,8]. Low health literacy has been shown to be associated with various poor health outcomes (eg, more hospitalization and higher mortality rate) and poorer use of health care services (eg, poorer ability to interpret health messages and take medications appropriately) [9]. People with low health literacy have different approaches to learning health information; lower health literacy is associated with higher use and more trust in health information from television, social media, blogs, or celebrity web pages as well as lower use of medical websites and less trust in health information from specialist doctors [10]. About 35% of the US population has only a basic or below basic health knowledge and is disproportionately represented by low-income or ethnic minority populations [11]. The ability of chatbots to provide comprehensible information to those with lower health literacy is one potential remedy for this inequitable health information disconnect.

The potential benefits that chatbots can provide have led to their implementation in a variety of health care contexts, including diet recommendations [12], smoking cessation [13], and cognitive behavior therapy [14], but more research needs to be done to understand how chatbots should be designed to be most effective. In a retail setting, it has been shown that a chatbot’s language and communication style influences ease of use and engagement [15]. However, users interacting with health care information chatbots may have different needs and expectations than when interacting with chatbots in other industries, and there is little research investigating the influence of design considerations of chatbots on their effectiveness for providing health care information. As chatbots have a history of being biased and unfair [16,17], efforts to explore design considerations of chatbots must account for the intersectionality of identities and be considerate of all people. A potential avenue for helping users connect with chatbots is to give the chatbot an identity or persona. It has been shown that other virtual agents may be more or less effective due to their perceived character [18], yet the effect that different personas have on the effectiveness of a health care information chatbot is unclear.

Thus, the primary objective of this study was to examine the effects of an educational health care chatbot, as it differs in complexity of responses (technical vs nontechnical language) and the presented qualifications of its persona (eg, Doctor, Nurse, or Nursing Student persona) on perceived usability, trust, and effectiveness. The secondary objective was to identify similarities and differences in how users conversed with the chatbot.

### Methods

#### Study Design

In this study, participants were tasked with interacting with the chatbot to seek information about blood pressure. The experiment was a 2×3 between-subjects design, in which the chatbot with which the participants interacted differed in the complexity of its responses (either technical or nontechnical language) and the presented qualifications of its persona (either Doctor, Nurse, or Nursing Student).

#### Ethics Approval

This study was reviewed and approved by the Clemson University Institutional Review Board (IRB2019-411).

#### Chatbot Design

The most common purpose of chatbots in health care has been to provide education and training for conditions (eg, mental health, type 2 diabetes, breast cancer, hypertension, asthma, pain monitoring, and language impairment) [19]. To emulate this common purpose, the chatbot created in this study was designed to answer questions and provide general health information about blood pressure. The chatbot used in this research emulated a pattern-matching chatbot rather than one which uses artificial intelligence. Pattern matching occurs when the question patterns match certain answer patterns. For this study, we created predefined answers that offered the same information in either technical or nontechnical language. The experimenter delivered the chatbot responses to questions asked by the participant using a “Wizard of Oz” technique. In this type of experiment, a participant interacts with a system that they expect to be autonomous but is secretly controlled by a member of the research team [20-22]. A prepopulated response list to possible participants’ questions was created, evaluated, and refined through pilot testing. The responses were created to address all questions that pilot tests identified as well as other possible generic question responses. These generic responses accounted for unanticipated questions or off-topic discussions not related to blood pressure. The generic responses did not change between technical and nontechnical conditions. An example of a generic response is, “I am sorry, I am unable to answer that question. Do you have another question about blood pressure?” An intensive care unit nurse was consulted to verify our chatbot content and to identify any additional information we may have missed or that was outdated or incorrect.

To differentiate between the complexity of the responses (technical vs nontechnical), we assessed the reading difficulty of each chatbot response using the Microsoft Word Reading Assessment feature. This feature uses the Flesch-Kincaid readability test, which determines a text’s Flesch reading ease and its Flesch-Kincaid grade level. The Flesch-Kincaid
assessments have been used to assess technical manuals, legal documents, and insurance policies [23,24]. The nontechnical responses all had high reading ease and a reading grade level of 8 or below, whereas the technical responses had low reading ease and grade levels of 12 or higher. These reading grade levels were chosen because patient education materials have been found to have mean reading grade levels around 11-14, whereas recommendations for appropriate reading grade levels are 6-8 [25]. Although one possibility to increase the reading level of a response could have been to add additional text or information, this was not done to ensure consistency in the amount of information presented by the chatbot to the participants between technical and nontechnical responses.

The persona that the chatbot represented consisted of 3 possible naming structures (ie, Doctor, Nurse, or Nursing Student). Each of the chatbot personas were named Sarah with only the salutation changing between the conditions (eg, “Dr Sarah,” “Nurse Sarah,” or “Nursing Student Sarah”). This was done to avoid any implicit bias in the persona based on using different names. Each of the personas introduced themselves at the start of the chatbot engagement. For example, “Hello, my name is Dr Sarah. I’m here to help you learn about blood pressure today. You can ask questions about understanding blood pressure, learning how to manage or prevent high blood pressure, who is affected, and more. What is your first question?” Following the initial engagement, the persona identifier was used as an identifier in each response to the participant.

Participants
Participants were recruited from Clemson university; they were required to be between the ages of 18 and 26 years and to be able to read, write, and speak in English. Participants received a compensation of US $10 for 30 minutes of their time at the end of the session. Participants between the ages of 18 and 26 years were chosen so that the participant population likely had a similar (nominal) level of knowledge about blood pressure.

Procedure
Following informed consent procedures, participants completed a demographic survey and then an experimenter assessed the participants’ health literacy using the Short Assessment of Health Literacy—English [26]. Participants then completed a multiple-choice test on blood pressure topics (henceforth referred to as the “pretest”). The blood pressure topics included the effects of high and low blood pressure, factors associated with blood pressure issues, and risk factors for high blood pressure. These factors were included based on the content in health textbooks and web-based resources that discuss blood pressure, common questions, and common misconceptions [27-30]. After the pretest, participants were instructed on how to begin using the chatbot and were informed that they had up to 15 minutes to learn about blood pressure by interacting with the chatbot. The experimenter, stationed in a separate room from the participant, ran the chatbot using a Wizard of Oz type of structure (ie, they responded to the participants’ questions with preconstructed answers). After interacting with the chatbot, participants took the same multiple-choice test on blood pressure topics (henceforth known as the “posttest”). Following the posttest, participants were given the Post–Study System Usability Questionnaire (PSSUQ) [31] and a survey assessing the trustworthiness, credibility, and perceived ease of use of the chatbot [32].

Analysis
Participants’ perceived usability of the chatbot was measured via the PSSUQ [31] and was evaluated using a linear regression model. Participant’s trust in the chatbot was measured via a question asking how much the participant agreed with the statement “I trust the chatbot” on a 7-point Likert scale (“strongly disagree” to “strongly agree”). This Likert scale was converted to a binary variable representing those who trusted the chatbot (ie, participants that responded with “somewhat agree,” “agree,” and “strongly agree”) and those who did not trust the chatbot (ie, all other responses). Trust was evaluated using a binary logistic regression model. The chatbot’s effectiveness was operationalized as the difference in pretest versus posttest scores from the blood pressure knowledge test. Effectiveness was evaluated using a median split binary logistic regression model. All regression models started by including response complexity and chatbot persona as well as the following demographic variables: self-identified gender, health literacy, ethnicity, and student status (eg, graduate or undergraduate student). Demographic variables were removed from the model stepwise following Akaike information criterion minimization until a final model was reached. Additionally, a qualitative transcript review of the participants’ conversation with the chatbot was conducted.

Results
Descriptive Statistics
Initially, 74 students participated in the study; however, 3 participant’s data were removed from the data analysis—two due to incomplete data collection and one because the participant did not engage in the task (eg, not asking blood pressure–related questions throughout the experiment). Of the remaining 71 participants, 43 (60.6%) self-identified as female, 30 (42.3%) were graduate students, and 41 (57.7%) were undergraduate students. The average age of the participants was 21.87 (SD 2.58) years. The demographic results are presented in Table 1.
Table 1. Characteristics of study participants (N=74).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>21.87 (2.58)</td>
</tr>
<tr>
<td>Gender, n (%)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>28 (39.4)</td>
</tr>
<tr>
<td>Female</td>
<td>43 (60.6)</td>
</tr>
<tr>
<td>Race, n (%)</td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>49 (69)</td>
</tr>
<tr>
<td>African American</td>
<td>8 (11.3)</td>
</tr>
<tr>
<td>Asian</td>
<td>14 (19.7)</td>
</tr>
<tr>
<td>Student status, n (%)</td>
<td></td>
</tr>
<tr>
<td>Undergraduate</td>
<td>41 (57.7)</td>
</tr>
<tr>
<td>Graduate</td>
<td>30 (42.3)</td>
</tr>
</tbody>
</table>

Usability

The average usability score was relatively high (mean 6.00, SD 0.63), indicating high perceived usability of the system. A linear model was constructed to model the usability scores from the independent factors and resulted in residuals that were significantly skewed (Shapiro-Wilk test: skewness 0.959; \( P = .02 \)). Therefore, the PSSUQ average scores were transformed using a square transformation, resulting in a model with residuals that were identified as not being significantly skewed (skewness 0.976; \( P = .18 \)). The linear regression model (Table 2) revealed that participants who self-identified as males \( (P=.049) \) and participants who interacted with the “Nursing Student” persona of the chatbot \( (P=.02) \) were significantly more likely to report the chatbot as having a lower usability. Participants who were undergraduate students were significantly more likely to report the chatbot as having a higher usability \( (P=.03) \).

Table 2. Linear regression model predicting usability of the chatbot.

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Estimate</th>
<th>SE</th>
<th>( P ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>39.5</td>
<td>1.98</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Response complexity (technical language)</td>
<td>-2.34</td>
<td>1.59</td>
<td>.15</td>
</tr>
<tr>
<td>Chatbot persona (“Doctor”)</td>
<td>-3.32</td>
<td>1.97</td>
<td>.10</td>
</tr>
<tr>
<td>Chatbot persona (“Nursing Student”)</td>
<td>-4.52</td>
<td>1.96</td>
<td>.02</td>
</tr>
<tr>
<td>Gender (male)</td>
<td>-3.38</td>
<td>1.70</td>
<td>.049</td>
</tr>
<tr>
<td>Student status (undergraduate)</td>
<td>7.05</td>
<td>2.27</td>
<td>.03</td>
</tr>
</tbody>
</table>

Trust

Only 9 of 71 (12.7%) participants reported not trusting the chatbot. A binary logistic regression model predicting trust \( (Table 3) \) revealed that participants with higher health literacy were significantly more likely to trust the chatbot \( \text{OR} 2.04, 95\% \text{ CI} 1.11-4.00; \ P=.03 \). No other factors significantly impacted the reported trust in the chatbot.

Table 3. Binary logistic regression model predicting trust in the chatbot.

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>OR(^a) (95% CI)</th>
<th>( P ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>&lt;0.001 (&lt;0.001-1.51)</td>
<td>.07</td>
</tr>
<tr>
<td>Response complexity (technical language)</td>
<td>0.80 (0.17-3.58)</td>
<td>.77</td>
</tr>
<tr>
<td>Chatbot persona (“Doctor”)</td>
<td>0.86 (0.14-4.95)</td>
<td>.87</td>
</tr>
<tr>
<td>Chatbot persona (“Nursing Student”)</td>
<td>1.94 (0.27-17.8)</td>
<td>.52</td>
</tr>
<tr>
<td>Health literacy score</td>
<td>2.04 (1.11-4.00)</td>
<td>.03</td>
</tr>
</tbody>
</table>

\(^a\)OR: odds ratio.
Effectiveness

The median difference in pretest versus posttest scores was an improvement of 4 questions, and thus, a median split separated participants who had an improvement of 4 or more into a “high effectiveness” group (n=37) and those who had an improvement less than 4 into a “low effectiveness” group (n=34). A binary logistic regression predicting effectiveness (Table 4) revealed that participants who received technical language responses were significantly more likely to be in the high effectiveness group (OR 2.73, 95% CI 1.05-7.41; \( P = .04 \)) when compared to participants who received nontechnical language responses. No other factors significantly impacted the effectiveness of the chatbot.

Table 4. Binary logistic regression model predicting effectiveness of the chatbot.

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>OR(^a) (95% CI)</th>
<th>( P ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.87 (0.33-2.25)</td>
<td>.76</td>
</tr>
<tr>
<td>Response complexity (technical language)</td>
<td>2.73 (1.05-7.41)</td>
<td>.04</td>
</tr>
<tr>
<td>Chatbot persona (“Doctor”)</td>
<td>0.84 (0.25-2.72)</td>
<td>.76</td>
</tr>
<tr>
<td>Chatbot persona (“Nursing Student”)</td>
<td>0.52 (0.15-1.69)</td>
<td>.28</td>
</tr>
</tbody>
</table>

\(^a\)OR: odds ratio.

Qualitative Transcript Review

Analysis of the chatbot conversation transcripts reveals that all of the 71 participants followed the general knowledge–seeking task. However, there were elements of how participants interacted with the chatbot that varied. Only about half of the participants (35/71, 49.3%) asked at least one question using the singular “I” form, often concerning prevention for themselves (ie, “How can I prevent high blood pressure from occurring?”). Of these participants, most (25/35, 71.4%) asked more than one question using the singular “I” form. Generally, the “I” questions could be answered with generic responses, but occasionally participants would ask questions such as “Am I at risk?” which the chatbot, based on the current chatbot pattern matching structure, was not able to answer explicitly for each participant. Only one participant asked the chatbot about assisting others: “How can I help someone with high blood pressure?” When participants received an “I don’t know” response from the chatbot, they generally reverted back to general knowledge seeking with questions like “What is blood pressure?” or “Who is affected most?”

A handful of participants (5/71, 7%) used scenarios at some point in their dialogue to learn about specific factors that could put them at risk of high blood pressure. The scenarios were generally self-centric, in that the participants wanted to know if their specific life circumstances or choices could affect their blood pressure. Textbox 1 summarizes the scenario style questions from the transcripts that demonstrate these scenarios or concerns.

Textbox 1. Scenario quotes from chatbot transcripts.

**Quotes**

- I am 25 year old [sic] and my mother and father both have high blood pressure. What are the odds that I get high blood pressure?
- What if I work out but eat unhealthy [sic]
- For a young woman age [sic] 18, what is the likelihood of developing high blood pressure?
- Has [sic] stress in college aged kids started an increase in hypertension in younger people [sic]
**Discussion**

**Principal Results**

Chatbots are growing in use across the internet, not only for consumer products and websites but also within health care settings. This paper described an exploratory study investigating how the design of a chatbot might impact its perceived trust, usability, and effectiveness in a health information search setting. The chatbot’s language was based on previous health care research that demonstrated that patients’ understanding of health information changes with language style and structure [4,18] as well as the cost of low health literacy on the health care system annually [7,8]. Chatbot persona was studied because it has been shown that other virtual agents may be more or less effective due to their perceived character [18]. Our results found that the chatbot’s responses which used technical language significantly increased the chatbot’s effectiveness but had no impact on trust or usability. The chatbot persona used in this study was found to significantly impact usability but had no impact on effectiveness or trust. Additionally, participants with higher health literacy reported higher trust in the chatbot. This finding is consistent with health literacy literature, which finds that people with higher health literacy generally have higher trust [33,34]. The qualitative transcript review revealed interesting insights about how people may use chatbots to gather health information and what they expect chatbots to be able to understand. The variation in sentence structure and grammar may be indicative of different subsets of users who interact with the chatbot, though that was not examined in this study. The use of shorthand is particularly interesting because it resembles more of a general, all-encompassing search pattern rather than a directed question-asking search pattern, perhaps indicative of those participants viewing the chatbot not as a person (as the persona looked to represent) but more as a search engine. Such generic searching demonstrates the need for chatbots to be able to process multiple kinds of search entries, whether it be formal input, shorthand, or all-encompassing search terms. These results show the potential that careful design may have on improving the effectiveness, usability, and trust in health care chatbots.

**Limitations and Future Work**

A key limitation was the relative homogeneity of the participants within this study; participants were of similar ages (18-26 years) and education levels. Although this age range was selected to support a more homogeneous group of possible participants without direct experiences and knowledge associated with blood pressure, this does limit the generalizability of the study. Technical language responses may have been more effective because all of the participants were college students with relatively high health literacy, and thus, simplifying the responses may only have served as a detriment. In other populations with lower health literacy, nontechnical language may be more effective. Future work should more closely reflect the wider population ages, experiences, and health literacies in evaluating the usefulness of chatbots in health care applications. Additionally, future work should evaluate how the users’ identities and their intersectionality influence their interactions with chatbots to account for potential cultural and other biases that may be implemented in a chatbot’s design.

Health literacy and its impacts on chatbot language, trust, and usability need to be further studied. This study found that health literacy had an impact on the trust in the chatbot, which was to be expected based on previous research [33]. However, this study found that health literacy did not have an impact on usability, which is inconsistent with previous research [34]. Future research should use qualitative measures, such as interviews, to investigate why relationships or lack of relationships, such as language and effectiveness, health literacy and trust, or health and usability, are transpiring.

Another limitation is the simple persona used in this chatbot. This persona was not found to significantly impact effectiveness or trust. This may be because the persona used in this study was simple, and therefore, potentially unengaging; it included only a name and title, it did not have a picture or other visual stimuli, and it did not engage in any personalized dialogue (eg, asking the participant questions). This is supported in the qualitative transcript review, which found that most participants did not acknowledge Sarah (the chatbot’s persona), and few responded to the greeting, addressed Sarah at some other point in the dialogue, or thanked Sarah. Overall, most of the participants did not appear to engage with Sarah beyond its use as a chatbot to deliver information, suggesting that some participants used the chatbot as more of a conventional search engine rather than a conversational agent. Future studies should examine other ways of representing personas to evaluate whether personas in general are useful in this context. Other representations could include additional visual stimuli like pictures or avatar images. As the representations transform into 3D or virtual agents, the required characteristics need to change as well and follow other design patterns [18,35]. Additionally, this study examined only differences in the qualifications of the chatbot’s persona; further work should examine how larger differences in the persona’s identity may improve the chatbot’s effectiveness, usability, and trust. Given that the low health literacy portion of the US population is disproportionately represented by low-income or ethnic minority populations [11], personas that better reflect these minorities may aid in improving the chatbot’s effectiveness for these underrepresented groups. There may also be other user interface design strategies that better facilitate the effectiveness of chatbots for these groups.

Neither language nor persona had a significant effect on trust in our study. This could be in part due to trust being difficult to measure and quantify [36,37]. Trust is complex and dynamic with multiple factors contributing to an individual’s trust [38]. It is also possible that the participants in our study developed negative trust or conditional trust, where individuals expected the chatbot to fail at some point (ie, negative trust) but still reported trusting it or expected that the chatbot could do certain things or tasks in certain contexts (eg, focusing only on blood pressure information from a health care chatbot) and still reported trusting it (ie, conditional trust) [36]. An example of the negative trust may have occurred when even the 9 participants who received 5 or more responses of “I don’t know” to their questions still had relatively high trust. Other studies have shown that using different relational strategies (eg, small talk and empathic reactions) was not able to foster trust in a chatbot [39].
Lastly, although the experimental setting attempted to replicate a health care website with a chatbot, the setting was a static website with a simulated chatbot. The responses were not truly determined by an artificial agent but were instead accomplished with preconstructed responses resembling a messenger type system via a Wizard of Oz study. This replication may have impacted the results, as the responses were simulated by an experimenter and not by the technology. Since the responses were given by a person, there is a possibility for variability in how the experimenter responded. Along with the experimenter’s possible variability, there was variability in what questions participants asked and how participants asked those questions.

Conclusions

With increased internet use in everyday life, the ways in which people obtain health care information are changing. It is important to continue to develop proper health care websites with information that can be personalized for users based on influential factors, such as age, gender, identity, and health literacy [5,8,40]. The ability of chatbots to provide personalized, private, and understandable health care information on a variety of topics makes it a promising tool, as health care trends toward web-based and remote processes. As participants look for health recommendations in different contexts and environments and with different devices and technologies, chatbots will need to be able to adapt to different needs. Understanding how those personal needs should change the language or presentation of the chatbot is crucial. Personalized health care information that is understood by each patient and caregiver will allow people to maintain ownership and have confidence in their health care decisions. As patients are better able to understand their health care needs, they can make decisions that allow for quicker recovery, create less impact on the health care system, and ultimately lower overall costs for the patient and the health care system.

Health care chatbots and telehealth medicine are also on the rise, not only in the last decade but particularly as a response to the COVID-19 pandemic. One technology implementation that saw an increase was telehealth medicine, where doctors and patients communicated virtually via videos, emails, and chats. Chatbots may be effective for these particular cases [41]. The COVID-19 pandemic additionally highlighted the global problem of health literacy disparity, as now more than ever people are forced to make health information–based decisions [42-44]. Therefore, an understanding of how to design and implement chatbots to effectively deliver health information is more crucial than ever. In order to develop effective design recommendations and guidelines for health care chatbots, future research needs to continue exploring how individuals perceive and interact with health care chatbots and their associated personas.

Conflicts of Interest

None declared.

References


Abbreviations

OR: odds ratio
PSSUQ: Post-Study System Usability Questionnaire

© Joshua Biro, Courtney Linder, David Neyens. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 01.02.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Development of a Secondary Prevention Smartphone App for Students With Unhealthy Alcohol Use: Results From a Qualitative Assessment

Nicolas Bertholet\textsuperscript{1,2}, MD, MSc; Elodie Schmutz\textsuperscript{1,2}, MSc; John A Cunningham\textsuperscript{3,4}, PhD; Jennifer McNeely\textsuperscript{5}, MD, MSc; Gerhard Gmel\textsuperscript{1,2}, PhD; Jean-Bernard Daeppen\textsuperscript{1,2}, MD; Véronique S Grazioli\textsuperscript{6}, PhD

\textsuperscript{1}Addiction Medicine Unit, Department of Psychiatry, Lausanne University Hospital, Lausanne, Switzerland
\textsuperscript{2}Faculty of Biology and Medicine, University of Lausanne, Lausanne, Switzerland
\textsuperscript{3}National Addiction Centre, Institute of Psychiatry, Psychology and Neuroscience, Kings College London, London, United Kingdom
\textsuperscript{4}Centre for Addiction and Mental Health, Toronto, ON, Canada
\textsuperscript{5}Section on Tobacco, Alcohol and Drug Use, Department of Population Health, New York University Grossman School of Medicine, New York City, NY, United States
\textsuperscript{6}Department of Vulnerabilities and Social Medicine, Center for Primary Care and Public Health, University of Lausanne, Lausanne, Switzerland

Corresponding Author:
Nicolas Bertholet, MD, MSc
Addiction Medicine Unit
Department of Psychiatry
Lausanne University Hospital
Bugnon 23 A
Lausanne, 1011
Switzerland
Phone: 41 0213148400
Email: Nicolas.Bertholet@chuv.ch

Abstract

Background: Despite considerable efforts devoted to the development of prevention interventions aiming at reducing unhealthy alcohol use in tertiary students, their delivery remains often challenging. Interventions including information technology are promising given their potential to reach large parts of the population.

Objective: This study aims to develop a secondary prevention smartphone app with an iterative qualitative design involving the target population.

Methods: The app development process included testing a first prototype and a second prototype, developed based on the results of 2 consecutive qualitative assessments. Participants (aged \textgreater{}=18 years, screened positive for unhealthy alcohol use) were students from 4 tertiary education institutions in the French-speaking part of Switzerland. Participants tested prototype 1 or prototype 2 or both and provided feedback in 1-to-1 semistructured interviews after 2-3 weeks of testing.

Results: The mean age of the participants was 23.3 years. A total of 9 students (4/9 female) tested prototype 1 and participated in qualitative interviews. A total of 11 students (6/11 female) tested prototype 2 (6 who tested prototype 1 and 5 new) and participated in semistructured interviews. Content analysis identified 6 main themes: “General Acceptance of the App,” “Importance of the Targeted and Relevant App Content,” “Importance of Credibility,” “Importance of the App Usability,” “Importance of a Simple and Attractive Design,” “Importance of Notifications to Ensure App Use over Time.” Besides a general acceptance of the app, these themes reflected participants’ recommendations toward increased usability; to improve the design; to include useful and rewarding contents; to make the app look serious and credible; and to add notifications to ensure its use over time. A total of 11 students tested prototype 2 (6 who tested prototype 1 and 5 new) and participated in semistructured interviews. The 6 same themes emerged from the analysis. Participants from phase 1 generally found the design and content of the app improved.

Conclusions: Students recommend prevention smartphone apps to be easy to use, useful, rewarding, serious, and credible. These findings may be important to consider when developing prevention smartphone apps to increase the likelihood of app use over time.

Trial Registration: ISRCTN registry 10007691; https://www.isrctn.com/ISRCTN10007691
Introduction

Unhealthy alcohol use is a leading cause of morbidity and mortality among young people, including among students in whom unhealthy alcohol use is associated with academic impairment, damage to self and others (assaults, unprotected sex, suicide, interpersonal violence), and institutional costs (property damage) [1]. Despite prevention efforts, consequences tend to increase over time [2]. In Switzerland, a significant proportion of the mortality among young people is attributed to alcohol [3,4].

The screening and brief intervention has demonstrated efficacy in primary care as an approach for nontreatment-seeking individuals [5-7]. Information technology has the potential to offer access to the screening and brief intervention to larger parts of the general population [8,9].

According to a 2019 Pew Research Center report [10], 76% of people in advanced economies are reporting smartphone ownership. The proportion of the population owning a smartphone is especially high for 18-34-year olds: 95% in the United States, 90% in Canada, 97% in France, 98% in Germany, 98% in Italy, 93% in the United Kingdom, 95% in Spain, and 99% in the Netherlands [10]. Given its widespread use, the smartphone may be an excellent tool to disseminate interventions, especially among young individuals. In a context in which there is a demand for electronic interventions [11], the development of smartphones offers an opportunity for more proactive interventions, with the potential for multiple contact at the user’s convenience, which may help increase the intensity of interventions.

Although the development of smartphone apps related to alcohol use has exploded, there is limited evidence regarding their efficacy to reduce unhealthy alcohol use [12,13]. The scarcity of evidence is particularly noticeable at a time when numerous apps are being developed and released. In addition, among the current apps focusing on alcohol reduction, few contain evidence-based behavior change techniques [14] and even thoughtfully developed apps can be associated with unanticipated adverse effects [15].

We are conducting a larger mixed methods study aiming to develop and test a smartphone app for unhealthy alcohol use among tertiary students through a randomized trial [16]. This paper presents the development of the smartphone app targeting unhealthy alcohol use. In this qualitative study, students were involved in an iterative process aiming at developing the app suitable to its target population. The developed app is currently being tested in a randomized trial.

Methods

Overview

We developed and tested in a previous pilot study a smartphone app targeting unhealthy alcohol use, based on a web-based intervention with demonstrated efficacy among young individuals [17,18]. We further developed this existing proactive secondary prevention smartphone app, taking into account the limitations observed during the pilot studies. The app was designed to offer additional features, taking advantage of the specificities of smartphones (ie, increased level of personalization and immediacy or access to intervention material in situations outside of the reach of face-to-face or computer interventions). As for numerous electronic interventions targeting unhealthy alcohol use, this app includes a social norms intervention [1,19-21], a type of intervention considered effective for college students [22]. The social norms intervention consists of normative feedback. The user’s alcohol consumption is compared with the alcohol consumption of people of the same age and sex in Switzerland, based on Swiss population data [23]. Normative feedback is provided for the volume of drinking (number of drinks per week) and for the frequency of heavy drinking episodes (frequency of episodes with ≥5 drinks [men]/≥4 drinks [women]). The app also provides personalized feedback on the risk of harm. Personalized feedback is considered one of the possible mechanisms of brief interventions to reduce alcohol use [24,25]. In addition, the app was designed to encourage self-efficacy through autonomous goal setting. It also provides additional information and resources to users willing or needing more: information and contact options for local addiction and mental health treatment resources, including student health centers, are listed.

The development of the app included the following iterative steps: (1) development of the initial prototype based on the app tested in pilot studies [26,27]; (2) test of the initial prototype in the target population and qualitative assessment; (3) app adjustments based on qualitative findings; (4) test of the second version (prototype 2) of the app in the target population and qualitative assessment; (5) final adjustments based on qualitative findings. Figure 1 shows the app development process.

The development of the initial prototype (ie, prototype 1 in Figure 1) was conducted through regular meetings between members of the research team and developers. Throughout the development process, informal testing was performed by members of the research team, developers, colleagues, and students working in our unit to ensure usability as well as test and design appropriateness. Notably, the aim was to make the app more “active” and to send messages to participants (ie, notifications) following prespecified scenarios, knowing that students appear acceptant of receiving such messages [28] and
that smartphone app users identify prompts as important [29].

Two groups of students, members of the app’s target population, were recruited to test the app in 1 (prototype 1) or 2 (prototype 2) of 2 tests and participate in in-depth interviews about the content of the app.

Figure 1. The app development iterative process.

**Materials**

The initial version of the app content in prototype 1 comprised 5 modules: (1) personalized feedback on self-reported alcohol consumption, with normative feedback, feedback on the calorific content of the reported consumption, and feedback on health risks (hereafter referred as “quiz”); (2) the blood alcohol content (BAC) computation module (hereafter referred as “test”); (3) the goal-setting tool (hereafter referred as “challenge”) with the possibility of obtaining a “badge” if the goal is achieved; (4) the designated driver tool (hereafter referred as “driver”); and (5) fact sheets (hereafter referred as “pedia”). The content of the app followed the existing literature [14,30-34] and previous research involving electronic interventions conducted by our group [17,18,26,27,35,36].

**Participants and Procedures**

Participants were students from 4 tertiary education institutions in the French-speaking part of Switzerland (ie, higher education institutions in health: Haute Ecole de Santé Vaud [HESAV]; University of Lausanne [UNIL]; Federal Polytechnic School of Lausanne [EPFL]; and Lausanne School of Hotel Management [EHL]). To be eligible (in test 1 or 2 or both), participants had to be a student, fluent in French, own an iPhone (Apple Inc; although the final version of the app is available for both iOS and Android, the development was carried out on an iOS platform), score 4 or more (for men) or 3 or more (for women) in the AUDIT-C (Alcohol Use Disorder Identification Test – Consumption) [37,38], and provide informed consent to participate. Study promotion was conducted with the support of students’ associations or communication staff of the targeted institutions. Study presentation and contact information were displayed by flyers (displayed in the various institutions) and electronically (eg, on the students’ association page on Facebook). Interested students participated in a phone call aiming to provide information about study participation, answer all questions, and screen for eligibility. Eligible students who were willing to participate completed written informed consent and were then explained about how to download the app. As the app was in the development phase, it was not available to the public and was specifically installed on each individual smartphone. Participants then tested the app for 2-3 weeks and took part in a semistructured interview to provide feedback. During the testing period, participants received reminders to keep testing the app. Prototype 2 of the app was developed based on feedback from test 1. Issues identified by multiple participants were given priority when refining the app.

Participants included in test 1 were invited to retest the app in test 2. In addition, new participants were included to ensure a naive feedback on the app. The inclusion criteria and procedures mirrored those used in test 1. Participants received CHF 50 (about US $50 at the time of the study) at the end of each
semistructured interview. All interviews were conducted by a senior researcher (VSG), in-person, using an interview guide (see Multimedia Appendix 1). The interviews aimed at exploring the following themes: general impressions on the app, perceptions regarding the app’s usefulness, perceptions on the app functioning, perceptions on the app design, perceptions regarding the app content, perceptions regarding each module of the app, and perceptions regarding the notifications. Each interview lasted between 36 and 70 minutes. Interviews were audio-recorded and transcribed verbatim by trained research assistants. All names and identifying information were removed before analyses.

Qualitative Data Analysis

Qualitative data were subjected to conventional content analysis. This method enables description of qualitative data through a systematic process of coding and classification [39]. Qualitative data were reviewed by VSG to identify recurring categories. Initial coding was conducted using a line-by-line technique aiming to narrate the actions occurring in the interviews [40]. After the initial coding process was completed, a codebook was created, wherein incident-by-incident codes were pooled and idiosyncratic or redundant codes were collapsed or eliminated. After the codebook was created, NB reviewed the codebook and tested it with 2 transcriptions. The codebook was then refined and finalized in consensus meetings. Finally, VSG rated all qualitative data from test 1 with the final codebook. The codebook was then adapted to rate qualitative data from test 2. Specifically, VSG reviewed a subset of interviews to identify new categories that did not appear in the original codebook. Finally, VSG rated all data from test 2. We present quotes from both tests 1 and 2 for illustration of categories emerging from the analysis. ATLAS.ti 7 (Scientific Software Development GmbH) was used to code qualitative data.

Ethics Approval

All procedures were approved by the local Ethics Committee (Commission cantonale d’éthique de la recherche sur l’être humain [CER-VD]; protocol number 2018–00560).

Results

Test 1

Multimedia Appendix 2 shows screenshots (by module) taken at the various stages of the app development process (prototype 1, prototype 2, and final version).

Participants

In total, 10 students were included in test 1. Of those, despite several reminders, 1 participant did not download the app and was therefore excluded, resulting in a sample of 9 participants. The latter tested the app for at least 2 weeks and took part in a semistructured interview. The mean age of the participants was 23.11 (SD 3.76) and 44% (4/9) were female. Qualitative results identified 6 main themes as described below. See Textbox 1 for a summary of the main themes and subthemes.
Textbox 1. Thematic framework summarizing the major themes and subthemes.

<table>
<thead>
<tr>
<th>Test 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Theme 1: General Acceptance of the App</strong></td>
</tr>
<tr>
<td>• General perceptions of the app</td>
</tr>
<tr>
<td>• App use in the future and recommendations to peers</td>
</tr>
<tr>
<td><strong>Theme 2: Importance of the Targeted and Relevant App Content</strong></td>
</tr>
<tr>
<td>• Personalized feedback on self-reported alcohol consumption</td>
</tr>
<tr>
<td>• The blood alcohol content computation module</td>
</tr>
<tr>
<td>• The designated driver tool</td>
</tr>
<tr>
<td>• The goal-setting tool</td>
</tr>
<tr>
<td>• The fact sheets module</td>
</tr>
<tr>
<td>• Add a monitoring tool</td>
</tr>
<tr>
<td><strong>Theme 3: Importance of Credibility</strong></td>
</tr>
<tr>
<td>• The cartoon character discredits the app</td>
</tr>
<tr>
<td>• The app must provide precise and valid results</td>
</tr>
<tr>
<td><strong>Theme 4: Importance of the App Usability</strong></td>
</tr>
<tr>
<td><strong>Theme 5: Importance of a Simple and Attractive Design</strong></td>
</tr>
<tr>
<td><strong>Theme 6: Importance of Notifications to Ensure App Use over Time</strong></td>
</tr>
</tbody>
</table>

Test 2

All themes and subthemes were the same as in test 1 except for the additional following subthemes:

- **Theme 2: The monitoring tool (based on feedback in test 1)**
- **Theme 3: The app looks more serious (based on feedback in test 1)**

**General Acceptance of the App**

**General Perceptions of the App**

All participants endorsed positive perceptions of the app that was commonly described as “stimulating,” “practical,” “fun, light, and not too serious.” When describing the app, participant 2 noted: “It is not heavy. That’s what I like about it; it’s not too serious, too heavy to make you ashamed to do it.” The app was also often perceived as “interesting” and “comprehensive.” Participant 1 explained: “I found the app interesting because it brings together many things about alcohol in one single app. So there’s no need for more. I liked much the fact it was pretty comprehensive.” Furthermore, half of the participants evoked the potential impact of the app on their alcohol-related behaviors:

- *Every time I drank a glass of alcohol, I thought about it [the app], so I can tell it had a restrictive effect on my consumption (...). I think I did not get alcohol once or twice because I told myself “well here if you enter four times in the app it will not make it!”* [Participant 3]
- *I think it [the app] is good given that I was still able to control my consumption of alcohol. Not regarding how much I drink, but to realize a little about the level of alcohol in my blood during the period and especially if all of a sudden, I’m driving or not since I have a driving permanent license.* [Participant 5]

Despite a general positive perception of the app, a few participants also commented on some general aspects they did not like: 1 participant reported that he was not stimulated a lot by the app that was judged as “too discreet.” Others questioned the alignment between the app content and the targeted audience, considering that it would be better suited to individuals with alcohol problems or to younger populations:

- *As I felt it was more for people who thought they had an alcohol problem and wanted to evolve, rather than for people who...students for example who just need to be sensitized and see the impact that alcohol can have on them.* [Participant 7]
- *I haven't actually learned much about my consumption and the consequences, but I think for people who are younger...For example, at the age when you start drinking a little bit, you really don't realize it. I think it can be even more interesting because at my age, I'm 23 (...), I know the consequences.* [Participant 8]

**App Use in the Future and App Recommendation to Peers**

Most participants reported that they would still use the app after the study, most often to assess their blood alcohol level when
partying. A minority of participants mentioned, however, that the app was not stimulating enough to have them using it over time. Participants 3 disclosed: “Well I think if it was really like that all the time, I would have done the process one more week and then I would have removed the app,” whereas participant 1 said: “This lack of reminders, I found it was not motivating to keep using it in the long run.”

When asking participants whether they would be willing to recommend the app to their peers, most answered that they would do so because they found the app was “fun,” “interesting,” and “easy to use.” Furthermore, participants commonly mentioned that they would recommend the app to their peers to limit the risks associated with driving while being intoxicated:

> I could recommend it to them [participants’ friends], because I think it could be very useful for those who have their license, it would help them calculate their consumption (...) it’s interesting to see where they stand to know if they should not drink too much...well, less drinking the time they drive and all. [Participant 4]

Finally, a few participants were unsure as to whether they would recommend the app to their peers because they would not like endorsing “the moralizer role” among friends or being judged because of the use of an alcohol-related app.

**Importance of the Targeted and Relevant App Content**

**Overview of the Content**

All participants agreed on the importance to provide targeted and relevant content. Specifically, participants’ feedback on the app content consistently reflected the idea that it must be interesting, stimulating, useful, and directly beneficial to them. We describe below participants’ feedback reflecting these ideas for each module enclosed in the app.

**The Personalized Feedback on Self-Reported Alcohol Consumption**

Participants consistently mentioned appreciating the personalized feedback, commonly perceived as the most “useful,” “interesting,” “relevant,” and “impactful” part of the app. They reflected positively on the opportunity to get personalized feedback and compare one’s own drinking with their peers:

> I think it’s good, precisely those numbers, because that’s where we really realize what we’re consuming and that’s when we fill out the questionnaires, well, we don’t really realize if we didn’t have any feedback, so I found it very interesting and that’s the big positive point. [Participant 3]

> That’s really the first incentive for me to use the app for a little...so typically with these results, I’d say that 0.5 percent of women your age drink alcohol like you do and 99.5 percent of women drink less alcohol than you do...When you get results like that, or at least when you’re above average, it’s a clear incentive to...well, to do something, or at least to be aware of the risks of your alcohol consumption. [Participant 1]

Likewise, participants frequently found the feedback on the calorific content of the reported consumption “interesting,” “relevant,” and “impactful.” Participant 2 explained for instance:

> I never think about the fact that it’s very caloric and to give a number like that on a hamburger, it makes an impact because I’ve had several times to say to myself “well I’m not going to eat at the [name of fast food] because it’s not good” and finally to realize that in one evening I may have eaten two [name of fast food], I say to myself “yeah okay.”

Similarly, most participants qualified the feedback on health risks as “interesting,” which allowed to increase awareness. Participant 5 mentioned that this feedback “allows realizing the health risks of excessive consumption.” He went on explaining that he found this particularly interesting because “you don’t really realize the problems associated with alcohol, it’s a bit of a decriminalized drug.” A minority of participants disclosed, however, that they did not read the information that was considered “too serious.”

**The BAC Computation Module**

All participants mentioned that they appreciated the BAC computation module and some of them considered this as the most useful part of the app. Most participants reported that the BAC computation module was the tool of the app they used the most frequently, generally while partying, to estimate whether or not they were able to drive home. Participant 5 reported for instance “[he] found [this part] the most useful...maybe because [he] was driving (...) and at least [he] knew that [he] could be a little safer.”

Most participants perceived the information on the risks associated with BAC as “useful” and “interesting,” whereas a minority of participants reported skimming over the reading, considering this information as “not very meaningful nor striking.”

**The Designated Driver Tool**

Most participants reported liking the designated driver tool that was frequently qualified as “funny,” “nice,” “useful,” and “practical.” Participant 2 recounted his experience while testing it: “People had a lot of fun taking pictures and then we would look at our faces in the pictures...the idea of taking pictures of everyone is fun.” The designated driver tool was also considered as potentially resolving a recurrent problem when partying. Participant 8 disclosed:

> Well I think it’s always the big debate when partying where nobody wants to be the driver and then even if we say “ok, ok, each one in turn”. We don’t necessarily go out with the same people every time...if everyone get the app it’s like that and not otherwise...And then at least there’s no discussion (laughs).

Despite these positive perceptions, however, most participants explained that this tool did not apply to them specifically because they do not drive when partying or have already...
implemented alternative safe habits. For instance, participant 6 explained: “I just tried it to see what it was like, because when we go out, given that we must take a car anyway, well we decide before who will drive, well who get a car.” That said, participants commonly considered that the designated driver tool did belong to the app, considering that it might be useful in younger populations.

The Goal-Setting Tool

Half of the participants indicated they appreciated the goal-setting tool that was often perceived as useful to set up limits, decrease alcohol consumption, or raise awareness of one own consumption:

> I think it can be pretty good because you can set a goal (...) you can even realize how much we drink because you have a little bit more perspective in this situation and say “ah, this week we drank more than I thought I would”, it allows you to see yourself and to realize. [Participant 5]

Three participants mentioned, however, that they felt this goal-setting tool did not apply to them, perceiving it more tailored to individuals with alcohol problems. Participant 7 said for instance: “I did not really understand its usefulness (...). For me it was more for someone who think he has a problem and all of it and that he wants to achieve challenges.”

Almost half of the participants indicated they liked the badges that were perceived as “nice,” “funny,” or motivating to set up challenges. Three participants found the badges interesting without giving much importance to it, whereas another participant considered they were not useful.

The Fact Sheets Module

Most participants mentioned they appreciated the fact sheets module that was considered “interesting,” “useful,” and sometimes “comprehensive” or “easy to read.” Participant 4 said: “Yes, no I found it good. Additionally it is not too long, it’s good.” A few participants suggested adding topics (eg, alcohol and drugs) or making it shorter or funnier, whereas 3 participants did not like the fact sheets module, considering it as “too heavy” or not funny.

Add a Monitoring Module

Almost half of the participants suggested adding a module aiming at monitoring alcohol consumption over time, without the need for goal setting. According to the participants, such an addition would help increase awareness of alcohol consumption over time and could make the app more stimulating:

> I would appreciate, I don’t know, at the end of the week to know how much I could have drunk or maybe

where, when, maybe with whom (...). I think it might be interesting to enter this type of data. I might tend to leave it in this (current) format if I really wanted to install it. When in fact I think I would be quite assiduous... it would already be interesting to know the results over a month, I think. [Participant 3]

And I would have liked to have, in relation to the evolution over time, but over a longer period of time, a kind of graph that would follow us over time as well. I would also have found it interesting to see more like that because well, the app is good... Well, yeah, I couldn’t figure out what I was consuming more than that either. [Participant 7]

Importance of Credibility

The Cartoon Character Discredits the App

All participants made comments highlighting the importance to make the app look serious and credible. In prototype 1, a cartoon character was included to guide participants through the app (Figure 2), a feature that has been used successfully in other electronic interventions [41,42].

The most consistent comment was related to this character that was commonly perceived negatively. Only a minority of participants reported they liked the character that was described as “funny,” “nice,” and “lightening the app.” Most participants explained they did not approve of the character that was considered “childish,” not tailored to the content and population targeted by the app, and ultimately at risk to discredit the app:

> Now, I'm not a big fan of the drawings themselves (...) I find it a bit childish, too childish. I find the idea of using character makes it lighter, but it makes it too light and I feel like it's a child's application when it's not at all and it's almost out of context (...) it looks like they're taken from a child's comic book and I don't find it logical. It makes the application a bit less credible. [Participant 2]

After that, it's... a small criticism, it's also the logo. I don't know if it's tailored to a consumption... it's a little bit very childish. (...) and isn't there a refusal to use an app where the logo is too childish because young people are not taken seriously. Maybe in their minds it's a bit of a refusal. [Participant 5]

Relatedly, participants commonly described the icon of the app (ie, displaying the face of the character) as “childish” or even “ridiculous.” Consequently, most participants recommended removing this character from the app and the icon.

Figure 2. The abandoned character included in prototype 1.
The App Must Provide Precise and Valid Results

Almost half of the participants questioned the normative feedback results’ validity, assuming that they were computed with data gathered by the app over time. Using data from serious research was perceived as necessary to consider and value the normative feedback results. When providing feedback on the normative feedback results, participant 2 explained: “I thought it wasn’t worth much because it came from...in comparison with other people using the app (...), whereas if I know that it comes from serious statistics, it will worth more to me.” After receiving an explanation related to the data used to compute the normative feedback, participants consistently recommended making this more visible in the app to increase the perceived normative feedback results’ value among future users (e.g., adding a pop-up after data completion or highlighting the research reference used to compute the normative feedback).

Furthermore, participants commonly questioned data taken into account to compute the BAC, reflecting a common expectation for accuracy and precision. Most participants mentioned that the BAC computation was approximate considering other influencing factors (e.g., precise number of consumption hours, having eaten or not) were not accounted for.

Relatedly, entering alcohol use data with the use of standard drinks was perceived as “too vague.” Participant 7 explained, for instance, that he found the questionnaire complex to fill in correctly because of the alcohol concentration differences across beverages:

> Filling in 100% right was complicated. (...) I like Belgian beer, which is often a bit strong (...) it’s a beer with more than eight percent, it’s not the same thing (as a regular beer). If I fill beer I’m not in the right category. So... and it’s the same at that point it’s also a bit complicated because what do you do? I still have my blood alcohol level, which is still double that of a standard beer.

Importance of the App Usability

All participants agreed on the fact that the app was globally easy to use. They commonly qualified the app as “clear” and “intuitive,” which was outlined as an essential ingredient to be further used. The fact that the app did not require time nor reflection to use was also commonly highlighted as positive and important. Participant 1 said: “It is simple and straightforward, no questions, and this is really good. In that, well, the easier, the more effective, the better.” Relatedly, participant 6 appreciated its interface because it was easy to use:

> You can see very well what it is used for. As you open it, you can see that you have the different options between the Quiz, to choose your driver, whether it’s to get information. So that’s good, I mean it’s visible and it’s clear what we can do.

In parallel, participants demonstrated very little perseverance when facing use challenges. Most difficulties involved entering alcohol consumption data in standard drinks. Participants commonly mentioned that they were unsure when filling in the alcohol questionnaire, most often because they hesitated about which category fits their drinks best. To address this difficulty, participants suggested adding information describing the categories and making it obvious to find. Participant 2 suggested: “Putting the ‘i’ [relating to information] near the alcohol categories to really explain what these categories correspond to, well to put the ‘i’ there because that’s where I would have looked for it.” Finally, a few participants reported difficulties to understand their normative feedback and suggested to make the feedback more visual and to simplify the sentences.

Importance of a Simple and Attractive Design

Only a minority of participants mentioned they liked the design, evoking the fact that it was not too serious or that the colors fit well. By contrast, participants commonly disliked the design, commenting it as “obsolete” and “not visual enough.” Similarly, the icon was frequently perceived as “nonvisual,” “not nice,” or “nonfinished.” Participant 1 said, for instance, that he “had the feeling to see a 2 or 3-year old app regarding the interface,” whereas participant 9 mentioned that “he found the lay-out not friendly” and not achieved, assuming it was related to the fact that the app was not over yet.

Consequently, participants recommended improving the design to make it more attractive while making it simple and sober. Participant 8 suggested for instance: “Well it could be a little more visual. I don’t know with different colors (...) or maybe a gauge with...if you’re in a non-risk situation in green and a risk situation in red,” whereas participant 5 disclosed: “It can be just something pretty elegant, simple, like a university campus app.”

Finally, similar comments applied to the badges included in the goal-setting module. Participants commonly disliked their design, considering they were “childish,” “too complicated,” and “not clear.” Participants explained they did not understand the meaning of the badges while looking at them, nor the association between the drawing and the meaning of the badge. Therefore, they recommended using more simple drawings and displaying them in a logical order to ease the understanding and make this part funnier and more stimulating.

Importance of Notifications to Ensure App Use Over Time

Participants generally mentioned that they found the notification frequency acceptable, although a minority suggested increasing its number. Participant 3 expressed being surprised to receive few notifications, which could have decreased the app use over time: “I was expecting to be more stimulated by the app (...) so I could have forgotten about it (...) without being solicited so I could have put it aside.”

Half of the participants considered that the notifications they received while testing the app were useful to remind them using the app and decrease the risk to forget the app:

> So I thought it was nice because sometimes it’s true that with the days that go by and everything is forgotten and sometimes I had the little notification and I thought “ah, I have to go and take a look at the app,” [Participant 4]

Participants expected receiving notification to remind them to fill in their ongoing goal-setting challenge. Participant 1

https://humanfactors.jmir.org/2023/1/e41088
explained he “did a 7-day goal-setting challenge and found surprising not receiving a reminder.” Hence, they recommended adding notification with specific content, most often to remind them fill in the ongoing goal-setting challenge. Corroborating previous comments, adding notification was perceived as essential to fill in data over time:

It’s true that these (goal-setting) challenges I forgot to fill them and then I gave up two or three of them because I forgot to fill them. Now it might be nice to get a notification that...so “today, don’t forget to fill in your consumption in the challenge.” [Participant 6]

**Modifications Based on Test 1 Qualitative Findings**

The major modifications concerned the design (change of name, icon/logo, home menu, general presentation, clarification of the alcohol use questionnaire with more specific options for standard drinks, presentation of feedback parts [normative feedback and BAC]) and the development of a new module allowing day-to-day monitoring. References for the normative feedback were added or placed in more prominent display to support the app’s legitimacy. A more readily available and more detailed description of standard drinks was included in the app with an option to access more detailed information on standard drinks. Different drinks choices were added. For example, an option for stronger beers/large beers (equivalent to 2 standard drinks) was added. The cartoon character was removed from the app’s icon. Similarly, the cartoon character that was created to help users navigate the app was removed. Within-app notifications and connecting messages (via within app pop-up messages or messages at the end of the feedback) between modules were included (ie, prompt to set drinking goals after an assessment of one’s alcohol use, indications on the evolution [increase or decrease] of one’s alcohol use since last use). How to obtain the different badges in the goal-setting module was also added (available when users click on any given badge). Table 1 presents how these badges can be earned. The full series of potentially collectible badges was displayed (with to-be-obtained badges grayed). Notifications and visual prompts were modified (notably, a red dot announcing that one should report something).

Table 1. List of the 15 badges and how they are earned.

<table>
<thead>
<tr>
<th>Badge type</th>
<th>Challenge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Badge 1</td>
<td>For the first completed challenge</td>
</tr>
<tr>
<td>Badge 2</td>
<td>For drinking exactly the amount of the goal (set by the user)</td>
</tr>
<tr>
<td>Badge 3</td>
<td>For drinking less than the goal on any challenge</td>
</tr>
<tr>
<td>Badge 4</td>
<td>For drinking half the amount (or less) of the goal</td>
</tr>
<tr>
<td>Badge 5</td>
<td>For drinking no alcohol during a 1- or 2-day challenge</td>
</tr>
<tr>
<td>Badge 6</td>
<td>For drinking no alcohol for 1 week</td>
</tr>
<tr>
<td>Badge 7</td>
<td>For completing 2 challenges in a row and drinking less than half of the goal</td>
</tr>
<tr>
<td>Badge 8</td>
<td>For completing 2 challenges in a row with no drinking</td>
</tr>
<tr>
<td>Badge 9</td>
<td>For completing 3 challenges (equal consumption)</td>
</tr>
<tr>
<td>Badge 10</td>
<td>For completing 3 challenges (lower consumption)</td>
</tr>
<tr>
<td>Badge 11</td>
<td>For completing four 7-day challenges without any alcohol</td>
</tr>
<tr>
<td>Badge 12</td>
<td>For completing 5 challenges</td>
</tr>
<tr>
<td>Badge 13</td>
<td>For completing a challenge during Saint Patrick’s Day</td>
</tr>
<tr>
<td>Badge 14</td>
<td>For completing a challenge during Christmas</td>
</tr>
<tr>
<td>Badge 15</td>
<td>For completing a challenge during New Year’s Eve</td>
</tr>
</tbody>
</table>

*For each challenge, the goal (ie, drinking limits) and duration of the challenge are decided by the user. The goal can be over the recommended drinking limits.

**Test 2**

**Participants**

The 9 participants who took part in test 1 were invited to test prototype 2 of the app. Of those, 6 accepted, whereas 3 declined because they were not available at the moment of the test (ie, all 3 were temporarily outside of the country). In addition, 5 new students were included in test 2, resulting in a sample of 11 participants. A comparison of the qualitative inquiry at test 1 between the 6 participants who accepted to take part in test 2 and the 3 who declined revealed no distinct trends differentiating their perceptions of the app. Semistructured interviews were conducted in-person with interview guides (see Multimedia Appendices 3 and 4). The mean age was 32.1 (SD 4.38) and 55% (6/11) were female. Qualitative results mirrored the main themes that emerged in test 1. Each theme is described below.

**General Acceptance of the App**

**General Perceptions of the App**

All participants—from test 1 and the new ones—spontaneously evoked at least once liking the app that was commonly described as “fun,” “interesting,” “not too heavy nor complicated,” “not
moralizing,” “welcoming,” and “not boring.” Similarly, most participants from test 1 spontaneously mentioned they found the app improved, most often with regard to its design. Participant 2 said for instance:

“I’m glad to see that the comments were taken into account and that it was...On the one hand so much better, but I noticed it especially on two points when I opened on the design level, it was especially on this point that I had insisted a lot. And it’s really changed a lot and I think it’s good.”

As many as 4 out of the 5 new participants evoked the potential impact of the app on their alcohol-related behaviors. Participant 10 (new) disclosed for instance:

“I was at a restaurant and I hesitated to have a beer and I thought “Ah I’ll have to put it in the app, that’s silly!” And so I didn’t take it.”

App Use in the Future and App Recommendation to Peers
In line with findings from test 1, most new participants (3 out of 5) and those from test 1 reported they were willing to continue using the app after the study most often to help maintain awareness on their drinking. Participant 7 explained, for instance, that he “would keep using the monitoring to see how he drinks over a longer period of time.” In the same line, when asking new participants whether they would be willing to recommend the app to their peers, all answered they would do so generally because it could enable them being aware of their drinking or decrease alcohol-related risks when partying.

Importance of the Targeted and Relevant App Content

Content Overview
In line with results from test 1, all participants highlighted the importance of providing interesting, stimulating, and useful content.

The Personalized Feedback on Self-Reported Alcohol Consumption
Consistent with this idea, 4 out of the 5 new participants mentioned liking the personalized feedback that was perceived as “interesting,” “useful” to compare one’s alcohol use with peers, and even the “added value of the app” for some of them:

“It’s the first time I’ve seen it [the personalized feedback] in an application like this (...). I think it’s a very good idea. Because it gives a first feedback to say...with numbers, (...) it can either give a slap to say “Ah well look how you’re consuming”, or on the contrary, it can be “Ah ok, that’s fine.” [Participant 12, new]

Likewise, new participants commonly perceived the feedback on calorific content of the reported consumption as “interesting,” “fun,” and “punchy.”

The BAC Consumption Module and the Designated Driver Tool
In the same vein, 4 out of the 5 new participants mentioned that they appreciated these modules and the information on the risks associated with BAC that was considered “useful,” “interesting,” and “informative.” Participant 11 reported, for instance, that he found the BAC computation tool was useful to set up drinking limits when one drives.

The Goal-Setting Tool
As many as 3 of the 5 new participants reported they liked the goal-setting tool, most often because it was considered “useful.” Participant 13 disclosed for instance:

“I like it a lot and I thought it was great because it makes you realize how much you’re drinking and at the same time, you can say: “Yeah next week I’m trying to have one less drink, see how I feel, see if it’s going to change something at my party, change something in my body.”

On an interesting note, 2 participants (1 from test 1 and the other new) suggested to add a 1-month goal-setting challenge in the tool. In line with findings from test 1, however, 1 participant felt that this module was more tailored to individuals with alcohol problems instead of students and half of the new participants mentioned they did not use the goal-setting tool because they found it was not useful to them.

All new participants showed interest in the badges, perceived as “fun,” “motivating,” and “rewarding.” However, 2 of them questioned the choice of badges’ names, perceived as “meaningful” or not directly related to the badge content. Relatedly, using English words and referring to foreign holidays (eg, Saint Patrick’s Day) were questioned, highlighting the importance of relevant content tailored to the target population.

The Monitoring Tool
Most participants from test 1 and 4 out of the 5 new ones showed interest in the monitoring tool. Almost half of the participants mentioned that this tool was the best part of the app. The monitoring was perceived as “relevant,” “impactful,” “interesting,” and “useful.” Participants commonly evoked that they appreciated following their alcohol use over time. Likewise, the related statistics were perceived as “interesting,” “meaningful,” and “impactful”:

“Finally the section I preferred is the monitoring. To indicate your consumption without necessarily giving a constraint it just allows you to see it written and to see “Ah, today I drank so much, today I drank so much. I say to myself,” “Ah, maybe I drank a little too much.” [Participant 12]

And what I liked is that we could have the statistics. So that’s really good because you can see between weekends or between certain times when if all of a sudden we increase our consumption or if we decrease it or if we are a little stable, I thought that was pretty good. [Participant 4]

A few participants mentioned, however, that recording drinking is “constraining” and requires assiduity and that only motivated users might keep monitoring their alcohol use over time. To help address this issue, participants consistently recommended adding notifications.
The Fact Sheets Module

In line with results from test 1, new participants consistently mentioned they appreciated the fact sheets module perceived as “comprehensive,” “interesting,” and “easy to read.”

Importance of Credibility

The App Looks More Serious

As many as 4 of the 6 participants from test 1 spontaneously reported that they found the design improved because it was “less childish” and “more tailored to the target public.” The 2 other participants from test 1 described the design as “more serious” and “less fun”; they noted that they liked the light design from the previous version while finding the new one acceptable. Participant 3 disclosed for instance:

> I though it [the design] was already good because it wasn’t too serious. Now it’s a little more serious but it doesn’t bother me at all because you still need to be credible for the application. So yeah I think it was a good idea to change the name and logo.

The App Must Provide Precise and Valid Results

Similar to test 1, participants commonly highlighted the importance of precision and validity in the app content. Participant 3 explained, for instance, that “when using this kind of app he wants to make sure to get convincing and reliable findings.” The most common comment in this regard was related to the BAC computation module. One of the new participants commented that he used another calculator to verify the results, leading to slight differences, which was perceived as potentially “discrediting the app.” Relatedly, other participants (new and from test 1) suggested ways to improve precision of the results, such as “providing the possibility to indicate the precise time of consumption” (ie, times of the first and last drinks), or adding an option to account for food consumption.

Other comments questioning the precision of the findings were related to the questionnaire measuring alcohol consumption. Some participants suggested adding the possibility to indicate alcohol percentage included in the drinks instead of indicating the beverage itself to improve measurement precision. Furthermore, the item measuring typical alcohol use per day (ie, “on a typical day, how many drinks do you drink?”) resulted in understanding issues in half of the participants from tests 1 and 2. Participants commonly reported being unsure about how to answer this question correctly, potentially leading to unprecise answers. Participant 2 suggested improving the item to make it clearer:

> Does that mean that if I’m used to drinking (...) we’ll say 7 beers but only once a week, should I put that I drink 1 beer a day every day or should I put that on a drinking day I drink 7 beers. It’s really not clear to me. For me it’s the “usual” that’s wrong. (...) in my opinion it would make more sense to write down on a drinking day, how much alcohol do you drink? In any case, to make that more explicit.

Finally, unlike test 1, participants did not question the validity of the personalized feedback results. Some participants from test 1 noticed some improvements in this area in the new version:

> I was happy to see that you took the comment I made into account, that we didn’t know where those numbers came from. Because the first time I didn’t know whether it was in relation to other users of the application or in relation to statistics that came from somewhere. So now it’s clear. [Participant 2]

Importance of the App Usability

Echoing results from test 1, all participants perceived the app as globally easy to use, commonly described as “clear” or “intuitive,” which was perceived as important. For instance, referring to the BAC computation tool, participant 13 said: “It is something very quick to do. Hop, hop and tac and it is done, which I found good.” In the same vein, the monitoring tool was commonly perceived as “easy to use.” Three participants suggested linking the different modules enclosed in the app to avoid users filling in their alcohol consumptions several times. Participant 1 explained for instance: “if I take the time to fill up my blood alcohol level for example, or a challenge, it can be interesting that it is directly found in monitoring.”

In parallel and consistent with findings from test 1, participants demonstrated little perseverance when facing use difficulties and recommended to make things obvious and clear wherever possible. The goal-setting tool caused the most difficulties in use. Participants were commonly unsure about how to use it and confused regarding the meaning and use of badges:

> You answer but you don’t know which one...well, if you answer a challenge, if you validate a challenge, well, I don’t know which badge I get. Maybe it’s written underneath the badges, but I admit I didn’t make the effort. [Participant 3]

In response, participants recommended to make this module clearer by adding information to guide users more efficiently.

Importance of a Simple and Attractive Design

The perceptions of the design were globally more positive in test 2 than in test 1. As many as 4 of the 6 participants from test 1 mentioned that they found the design was much improved, qualified as “more mature” or “professional.” Similarly, perceptions of the design were positive among 4 of the 5 new participants who described it as “simple,” “clear,” or “efficient” (the fifth new participant perceived the design as “too basic”). Unlike in test 1, the modified icon of the app was endorsed by most participants who described it as “much better than in the previous version,” “clear,” or “fitting the app content.” Similarly, most participants appreciated the new version of the personalized feedback, which was commonly considered “easy to understand,” “visual,” and “clear.”

Although perceptions of the design were globally more positive than in test 1, participants made several recommendations to improve it. Participants (from test 1 and new ones) recommended, for instance, adding more colors in general and using pictures illustrating beverages in the alcohol questionnaire to make the design more attractive:
In the questionnaires and in the choices of the types of alcohol we drink (…) what could perhaps be better in my opinion and a little more fun would be to use pictures instead. Maybe a little more illustrations again. You see it’s really kind of something that people like, to have a visual aspect in this app, a little bit colorful and all that. (…) That it’s more eye catching. [Participant 14]

Similarly, almost half of the participants recommended making changes in the app home screen that was qualified as “unattractive” and “too basic.” Improvement suggestions included rounding the corners of rectangles, using icons instead of white bars, and again, using more colors and pictures.

Importance of Notifications to Ensure App Use Over Time

Consistent with results from test 1, participants expected receiving text notifications to remind them to fill in their ongoing goal-setting challenges. Similarly, participants commonly recommended adding text notifications to remind users to fill in the monitoring:

It was after I registered a challenge, I didn’t get a reminder. I had done a kind of diet, I had downloaded an app that asks for everything you eat (…) it would write to me every time after a meal “Don’t forget to tell us about the meal” And finally it’s true that I would say to myself “Ooh I forgot to fill in” and then I would go. well I like to be reminded because I tend to forget. [Participant 13]

I realized that for the monitoring, it is noted in red and I did not receive any notification (…). I almost wish I had a notification that said during the day “record your consumption” so that I would remember to do it. (…) Because it’s actually within the app and if I don’t go there every day, I don’t see it. Except that I don’t think about going there every day. It’s true that if you want to monitor, it would be nice to have a notification per day, since it’s a daily analysis. [Participant 7]

Modifications Based on Pretest 2 Qualitative Findings

Further modifications were made on the design with modifications to the home page (icons added), ordering of modules, graphical presentation of the normative feedback results, and presentation of the monitoring data. Data entered in the monitoring module were transferred into the goal-setting module (if activated) and vice versa. The structure of the fact sheets module was modified. A 1-month challenge option was added. Presentation of the BAC calculator results was updated with presentation of possible symptoms at the BAC reached for the reported consumption. Because of financial constraints and availability of numerous calculators, we did not include the possibility to report food intake as part of the BAC computation. Screenshots of the final app are presented in Multimedia Appendix 2.

Discussion

Principal Findings

We developed a smartphone app targeting unhealthy alcohol use for students using an iterative development process. This study indicates that a smartphone app is an acceptable way to deliver unhealthy alcohol use interventions to students, a population with notable technology skills. Although the app was generally well accepted and appeared to be a suitable mean to deliver a brief intervention for unhealthy alcohol use, qualitative interviews allowed us to identify important aspects for the target audience: the app has to have a high level of usability, its design must be simple and attractive, users must consider the app content targeted and relevant to their needs, the app should have a high level of credibility, and finally, notifications and prompts are crucial to keep users interested and engaged. Through the iterative process we were able to develop an app that incorporated evidence-based elements from other electronic interventions and that corresponds to the needs and perceptions of the targeted audience. Our results inform on app development and the need to focus on elements of relevance and scientific credibility, as well as clear and effective presentation of scientific data. In addition, independent of its content, an app has to present a high degree of usability (ie, any task required by the app that is not immediately understandable will be abandoned) and an up-to-date and targeted design. Any feedback must be easily and immediately interpretable (hence the necessity to present scientific data on readily understandable graphic format). The app should also be adapted to the local and cultural context and some features, such as a cartoon character, may be well received in some but not in other populations groups. Similarly, especially among students, some of whom may be studying computer programming and app development, there is a high sensitivity for an up-to-date design, which may limit the half-life of apps.

Currently, the evidence for efficacy of a smartphone app for unhealthy alcohol use is scarce and thus there is an urgent need for efficacy data. A key feature in being able to assess an intervention’s efficacy is giving the studied intervention the best chance to be used by its target population [43]. One of the challenges of electronic interventions is to have users involved with the intervention content, notably with repeated use of the app. This study adds that by involving members of the target population in the development of the intervention, substantial modifications can be made to the design and presentation of its components. These changes should lead to a more targeted app design. As an example, comments were made on the prototype 1 app icon which, for some users, would have led to a lack of interest in the app before even assessing its content. As such, design plays a crucial role in electronic interventions’ abilities to reach their target audience. While this qualitative study did not lead to changes in terms of the evidence-based components of the intervention, major changes were made on how the feedback was delivered, how results were presented, and how people would perceive the app, notably in terms of scientific credibility, noting that, in the studied population, the app developed by a university hospital was not sufficient and that data sources (eg, for the normative feedback) had to be
prominently displayed to increase credibility and relatedness to the feedback results.

**Limitations**

This study presents limitations. A notable limitation is linked to its own justification: while we targeted a population to develop a specifically suited intervention (ie, students from tertiary education institutions in Switzerland), the results are only relevant to this specific population and generalizability is limited. Nonetheless, broad categories are identified that are likely relevant to other populations as well. Other studies have shown the importance of credibility, ease of use, and tailored content [30,44-46] and while achieving this may differ according to different context and populations, we expect it to be relevant for other populations than just for the sample studied. As participants were specifically recruited to test the app, our study sample comprises people who were likely more motivated than the target population intended to use the smartphone app. Thus, results from this study will have to be compared with usage data that will be obtained in the second phase of the study (randomized trial).

Although our results are informative on the development process of the app, we currently do not have data on the app’s effectiveness in reducing unhealthy use of alcohol nor on how and when the app is used. This will be the subject of the ongoing randomized trial (started in May 2021).

**Comparison With Prior Work**

Our results bear similarities with other studies on electronic interventions. Notably, Baumel and Kane [30] have shown the importance of design efficiency in real-world usage of eHealth interventions. In a study on a smartphone intervention targeting drinking conducted among members of the British Armed Forces, Puddephat and colleagues [45] showed the importance of credibility of information, targeted and personalized content, ease of use, and simplicity [45]. Garnett and colleagues [31] also showed the importance of the credibility of scientific sources. In a usability testing study among young adults who participated in focus groups after using a smartphone app targeting harmful drinking, Milward and colleagues [44] showed the importance of design, ease of use, and tailored content. Outside of a research setting, available evidence indicates that users will choose an app based on its look and credibility and tailored content [47].

**Conclusions**

This qualitative study conducted among students shows that smartphone apps targeting unhealthy alcohol use need to have a simple and attractive design, tailored features, scientific credibility, be easy to use, and that the app should regularly send notifications.

**Acknowledgments**

The trial is funded by the Swiss National Science Foundation (grant 176295; [48]).

**Conflicts of Interest**

None declared.

---

**Multimedia Appendix 1**

Interview grid, test 1.
[DOCX File, 29 KB - humanfactors_v10i1e41088_app1.docx]

**Multimedia Appendix 2**

Screenshots, by module, of the app at the various stages of the development process.
[DOCX File, 6116 KB - humanfactors_v10i1e41088_app2.docx]

**Multimedia Appendix 3**

Interview grid, pretest interview 2, group 1 (pretest participants 1).
[DOCX File, 29 KB - humanfactors_v10i1e41088_app3.docx]

**Multimedia Appendix 4**

Interview grid, pretest interview 2, group 2 (new participants).
[DOCX File, 29 KB - humanfactors_v10i1e41088_app4.docx]

**References**


42. Braciszewski JM, Tzilos Wernette GK, Moore RS, Bock BC, Stout RL, Chamberlain P. A Pilot Randomized Controlled Trial of a Technology-Based Substance Use Intervention for Youth Exiting Foster Care. Child Youth Serv Rev 2018 Nov;94:466-476 [FREE Full text] [doi: 10.1016/j.childyouth.2018.08.011] [Medline: 31435121]


48. SNSF Data Portal. URL: http://p3.snf.ch/Project-176295 [accessed 2023-02-22]

**Abbreviations**

- **AUDIT-C**: Alcohol Use Disorder Identification Test – Consumption
- **BAC**: blood alcohol content
- **EHL**: Lausanne School of Hotel Management
- **EPFL**: Federal Polytechnic School of Lausanne
- **HESAV**: Haute Ecole de Santé Vaud
- **UNIL**: University of Lausanne

©Nicolas Bertholet, Elodie Schmutz, John A Cunningham, Jennifer McNeely, Gerhard Gmel, Jean-Bernard Daeppen, Véronique S Grazioli. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 07.03.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Feasibility and Acceptability of the Aboriginal and Islander Mental Health Initiative for Youth App: Nonrandomized Pilot With First Nations Young People

Kylie M Dingwall¹, BA (Hon), PhD; Josie Povey², BOT, MPH, PhD; Michelle Sweet³, PhD; Jaylene Friel², CertII (Community Research), CertIII (Business); Fiona Shand¹, PhD; Nickelai Titov⁵, PhD; Julia Wormer⁶, BSW, GradCert (Youth Mental Health); Tumoor Mirza⁶, MBBS; Tricia Nagel², MBBS, PhD

¹Menzies School of Health Research, Charles Darwin University, Alice Springs, Australia
²Menzies School of Health Research, Charles Darwin University, Darwin, Australia
³Menzies School of Health Research, Charles Darwin University, Adelaide, Australia
⁴Black Dog Institute, University of New South Wales, Sydney, Australia
⁵MindSpot Clinic, Macquarie University, Sydney, Australia
⁶headspace Darwin, Anglicare NT, Darwin, Australia

Corresponding Author:
Kylie M Dingwall, BA (Hon), PhD
Menzies School of Health Research
Charles Darwin University
PO Box 795
Alice Springs, 0871
Australia
Phone: 61 889595380
Email: kylie.dingwall@menzies.edu.au

Abstract

Background: Despite young First Nations Australians being typically healthy, happy, and connected to family and culture, high rates of emotional distress, suicide, and self-harm are also observed. Differing worldviews of service providers and First Nations young people regarding illness and treatment practices, language differences, culturally inappropriate service models, geographical remoteness, and stigma can all inhibit access to appropriate mental health support. Mental health treatments delivered digitally (digital mental health; dMH) offer flexible access to evidence-based, nonstigmatizing, low-cost treatment and early intervention on a broad scale. There is a rapidly growing use and acceptance of these technologies among young First Nations people.

Objective: The objective was to assess the feasibility, acceptability, and use of the newly developed Aboriginal and Islander Mental Health Initiative for Youth (AIMhi-Y) app and determine the feasibility of study procedures in preparation for future assessments of effectiveness.

Methods: This was a nonrandomized pre-post study using mixed methods. First Nations young people aged 12-25 years who provided consent (with parental consent where appropriate) and possessed the ability to navigate a simple app with basic English literacy were included. Researchers conducted one face-to-face 20-minute session with participants to introduce and orient them to the AIMhi-Y app. The app integrates culturally adapted low-intensity cognitive behavioral therapy (CBT), psychoeducation, and mindfulness-based activities. Participants received supportive text messages weekly throughout the 4-week intervention period and completed assessments of psychological distress, depression, anxiety, substance misuse, help-seeking, service use, and parent-rated strengths and difficulties at baseline and 4 weeks. Qualitative interviews and rating scales were completed at 4 weeks to gain feedback on subjective experience, look and style, content, overall rating, check-ins, and involvement in the study. App use data were collected.

Results: Thirty young people (17 males and 13 females) aged between 12 and 18 (mean 14.0, SD 1.55) years were assessed at baseline and 4 weeks. Repeated measures 2-tailed t tests showed improvements in well-being measures that were statistically and clinically significant for psychological distress (Kessler Psychological Distress Scale, 10-item) and depressive symptoms (Patient Health Questionnaire, 2-item). Participants spent on average 37 minutes in the app. The app was rated positively, with
mean ratings of 4 out of 5 points (on scales of 1-5). Participants reported that they found the app easy to use, culturally relevant, and useful. The feasibility of the study was demonstrated with a 62% recruitment rate, a 90% retention rate, and high study acceptability ratings.

**Conclusions:** This study supports earlier research suggesting that dMH apps that are appropriately designed with and for the target populations are a feasible and acceptable means of lowering symptoms for mental health disorders among First Nations youth.

*(JMIR Hum Factors 2023;10:e40111)*  doi:10.2196/40111

**KEYWORDS**

digital mental health; First Nations; Indigenous; young people; feasibility study; digital health; mental health; depression; mHealth; mobile app; aboriginal; acceptability; youth

**Introduction**

Young First Nations Australians are typically healthy, happy, and connected to family and culture [1,2]. A 2017 national survey revealed 76% of young First Nations Australians reported feeling happy most or all of the time, 73% felt they were able to have a say on important issues within their family, and 59% had very positive or positive feelings about the future [2]. Nevertheless, rates of emotional distress are high, with around one-third of First Nations’ young people experiencing high to very high levels of distress, 2 in 5 experiencing mental health conditions, and suicide being a leading contributor to the disease burden [2]. Despite the high need, First Nations young people are particularly reluctant to seek help [2]. Constraints include stigma, discrimination, differing worldviews, language differences, cost, and transport issues [3,4]. Adolescence is a particularly prime time for intervention, as approximately 50% of mental illness emerges during that period [5]. Implementation of successful prevention efforts early is more likely to prevent the onset of mental disorders [5]. Timely access to culturally appropriate, effective prevention and treatment for First Nations young people is therefore imperative.

There is increasing recognition of the potential for digital mental health (dMH) interventions to overcome access issues and address the unique and significant mental health needs of First Nations people in rural and remote areas [6]. The potential benefits of digital health solutions, particularly those developed and implemented in a culturally informed way, could be substantial [7]. There are several ways dMH tools can be used. They can be used as a stand-alone treatment or augment existing treatments through blended care or fully supported treatments [8]. One particular benefit is that they can extend therapeutic activities such as assessments, monitoring, support, and interventions into real-world settings [9,10]. In the First Nations context, dMH tools developed in a culturally responsible and appropriate manner have shown great potential in supporting the development, maintenance, and strengthening of First Nations cultural identity [11].

While there are many apps available, there is limited evidence base for the majority of them. A recent study found over 1500 apps publicly available for addressing depressive symptoms; however, only 32 studies were presented in the academic literature [12]. For young First Nations Australians, and particularly those in remote regions, there are limited culturally specific dMH options. A recent review identified 3 individual-directed web-based apps available in the Australian First Nations context [7]. These were MindSpot, iBobby, and Aboriginal and Islander Mental Health Initiative’s (AIMhi) Stay Strong, each with some level of evidence for their acceptability and effectiveness [6,13-17].

In Australia, recent randomized controlled trials (RCTs) have promoted culturally adapted mental health apps as a feasible and acceptable means of lowering symptoms of mental illness for First Nations young people [16] and adults [14]. The iBobby app has shown promise for reducing risk factors for suicide among First Nations young people through the use of acceptance and commitment therapy, mindfulness, and self-soothing activities [15,16]. Similarly, the AIMhi Stay Strong app (and the paper-based Stay Strong Plan) is a culturally adapted motivational care planning (MCP) therapy that has demonstrated effectiveness in reducing psychological distress, depressive symptoms, and substance misuse among First Nations adults [14,18]. This therapy uses a holistic and strength-based approach consistent with First Nations conceptualizations of social and emotional well-being [19] and incorporates culturally adapted low-intensity cognitive behavioral therapy (CBT) and motivational interviewing elements. The Stay Strong therapy adopts an empowering, person-centered perspective, incorporating the 4 steps of identifying supportive people, strengths, worries, and setting goals for change.

While effectiveness is generally assessed through an RCT, evaluation of the feasibility, acceptability, and engagement with dMH interventions requires varied methodologies, and a combination of both subjective and objective criteria are suggested as appropriate for understanding user engagement [20]. For example, feasibility can be assessed through recruitment, retention, adherence, and completion rates. Acceptability and engagement are often measured through the extent of use (eg, app opens, minutes spent in the app), along with user-reported subjective views and experiences [21,22]. Given the speed at which technologies advance, the methods for evaluating dMH tools need to be flexible and proportionate to the tools’ complexity and the anticipated size of the effect on users’ mental health [21,22]. The time and resources required for planning and undertaking an RCT are not always appropriate for the assessment of dMH tools due to their rapid development and the accelerating time to the obsolescence of technological innovations [9,22,23]. There is a need for better-quality early evaluation of new digital health products, whether an initial summative evaluation of an established product to help decide...
whether it is worth adopting or a formative evaluation of an app during its development [22].

In 2020, three years of co-design workshops with First Nations youth resulted in the development of the first version of a new mental health app, the AIMhi for Youth (AIMhi-Y) app [24,25]. The participatory design process identified young First Nations Australians’ lived experiences of mental health and well-being and the dMH tool features preferred by young people and service providers. It also assessed the alignment of these preferences with recommendations from the scientific literature (including CBT, behavioral activation techniques, self-monitoring, notifications, gamification, etc) to design the new app [24,25]. The aim of this study was to assess the feasibility, acceptability, and use of the newly developed AIMhi-Y app and to determine study feasibility in preparation for future assessments of effectiveness.

**Methods**

**Study Design**

This is a nonrandomized pilot study using mixed methods to assess the newly developed AIMhi-Y app. The feasibility of conducting a larger-scale trial was tested using an uncontrolled single-group prepost design. The focus was on assessing the feasibility of the study methods, app implementation, app user engagement, and outcome measures, as well as gaining feedback on the barriers and acceptability of the app and study methods. The intervention period was 4 weeks. Data collection included prepost-delivery of outcome assessments at baseline and 4 weeks, researcher observation during intervention delivery, completion of qualitative exit interviews including app and study ratings at 4 weeks, and review of app use data (see Figure 1).

**Participants and Setting**

Participants were a convenience sample of young people referred to the study from participating services in Darwin, Northern Territory (NT), including Anglicare NT headspace, Stars Foundation, Clontarf Foundation, and the Council for Aboriginal Alcohol Program Services. Eligible participants were those identified by key contacts at each organization as likely to benefit from an introduction to a well-being app. They included First Nations young people aged 12-25 years with basic English literacy and the ability to navigate a simple app, for whom informed consent was gained from the individual and their parent or guardian.

**Governance**

An Indigenous Youth Reference Group (IYRG) involving 21 young people (86% female), 15-25 years old, residing in regional (n=11), remote (n=5), or very remote (n=5) settings, met (in...
person or via videoconference) 4 times over the study period. They provided advice regarding study procedures, the interpretation of data, and ideas for future app iterations. Meetings lasted 1.5-2 hours and were facilitated by a First Nations youth researcher and a nonindigenous project manager. Twelve IYRG members had been involved in the earlier co-design and development of the AIMhi-Y app [24,25].

An Expert Reference Group (ERG) met 3 times, gathering 10 youth service and research professionals (external to the Menzies research team; 4 of whom were First Nations Australians). The ERG assisted in refining study procedures and interpreting and disseminating findings in an advisory capacity.

The Menzies research team made key decisions with advice from the IYRG and ERG and carried out oversight of the project and the day-to-day tasks of the research. It included 3 First Nations youth researchers studying vocational education and training certificates in community health research, community services, or business, a senior cultural advisor and Larrakia Nation traditional owner, as well as a nonindigenous project manager and 3 nonindigenous senior researchers. The First Nations youth researchers were trained in study procedures and supported by the project manager with clinical mental health experience. Although involved throughout all stages of the study, the youth researchers played a particularly key role in refining study procedures and leading engagement, consent, and data collection with the young people.

Ethics Approval

Ethical approval was obtained from the Menzies School of Health Research Human Research Ethics Committee (HREC 2020-3851), including the Aboriginal subcommittee and the Northern Territory Department of Education Research Ethics Committee (reference number 16287). We endeavored to achieve Indigenous data sovereignty through the review and approval of study design by the Aboriginal Ethics Subcommittee and engagement with the First Nations research team and Reference Group members in the co-design of data collection tools, data collection, and interpretation of data. Participants could control the data they inputted into the app and choose to maintain or delete it postproject.

Referral and Screening

Eligible young people were referred through key staff at each organization. Young people’s parents or guardians (if under 16 years old or attending school) were contacted by a research officer via telephone or face-to-face to gain informed (verbal) consent prior to contacting and gaining informed consent from young people. At the suggestion of the IYRG, an additional “study agreement” was included to ensure participants were aware of the expectations of them while in the study (ie, to contact a support person if distressed and to use the app at some point in the following 4 weeks in their own time so they could provide feedback). Referred young people were screened prior to entering the study using the Kessler Psychological Distress Scale (K10) to test the feasibility of the screening process (for a later efficacy study, which would potentially exclude those likely to be well (K10 score <20). However, those with K10 scores <10 remained in this study as the main focus was to gather feasibility and acceptability data. K10 is sensitive to symptoms of both anxiety and depression and has been extensively used in state and national First Nations surveys [26,27]. To ensure safety, young people were also screened for imminent suicidal intent using the adapted PHQ-9 suicide item: “Have you been thinking about hurting yourself or killing yourself?”; as well as with follow-up questions: “Have you EVER, in your WHOLE LIFE, tried to kill yourself or made a suicide attempt?” (from the PHQ-9 modified for teens [28]) and “Have you been thinking about hurting yourself or killing yourself TODAY?” Those answering yes to either follow-up question were excluded from the study and referred to the immediate care of the person identified as the organization’s key study contact with a recommendation for urgent mental health assessment. At the request of the Anglicare NT headspace site, the exclusion criteria were relaxed at that site only to include individuals with suicidal intent and those with symptoms of early psychosis. The Anglicare NT headspace team was keen to offer the service to this broader client group and was confident that they were otherwise well supported within their service.

Participant Demographics and Flow

There were 56 young people referred (with 43 able to be contacted), of whom 35 consented and were screened to enter the study (see Figure 1). Of the 35 participants, 2 were excluded due to imminent suicidal intent and referred to the site contact for follow-up. Two were unable to be followed up on, and one withdrew. Data were analyzed from 30 participants (17 males and 13 females) aged between 12 and 18 (mean 14.0, SD 1.55) years. The majority (n=25) spoke English at home (one also spoke Wadjai, another also spoke Kriol), and 3 spoke Yolngu Matha at home (2 not reported). The majority usually resided in Darwin, while 2 resided in a remote community (3 were not reported).

Intervention

The smartphone-based AIMhi-Y app (version 1.0) integrates culturally adapted low-intensity CBT, psychoeducation, and mindfulness-based activities [24]. Participants begin by assisting fictional characters through a series of levels of a “quest,” aiming to become familiar with the content before beginning their own quest (ie, inputting their own information). Each quest presents 9-10 levels, which include the 4-step AIMhi Stay Strong MCP therapy. This involves the identification of (1) supportive people, (2) strengths, (3) worries, and (4) lifestyle changes or goals, which are interspersed with psychoeducational videos and games (as separate levels) [13,18,24,25]. Activities and information target both anxiety and low mood. For example, videos describe the app and getting started; the tree metaphor for growing strong; tips for how one of the characters uses mindfulness principles to find his calm; and 4 simple steps users might do to get through tough times, such as talking with a trusted person, doing more of the things that keep them strong, doing less of the things that take their strength away, or getting help from a health service. A summary page collates user or character information and presents progress. Mini-games were included in an attempt to promote relaxation, encourage real-world activities, and provide fun, engaging, and immersive sensory experiences [29]. For example, the bubble game invites
users to pop bubbles as they float across the page while calming music is played; the fishing game uses a similar concept to “catch” fish; and the animal game invites users to imagine they are on a bush walk and asks them to listen to animal sounds and identify the animal. The app is easy to use with simple and intuitive designs and nonclinical, youth-friendly language. Aboriginal language words are interspersed throughout the stories, relevant to specific characters, but the majority is in English given the large number of Aboriginal languages spoken in the region. Users can select from the options presented, edit them, or input their own text at each of the 4 steps. Vibrant colors and design elements reflect the natural landscapes of different NT regions (see Figure 2). The app can be used offline, but a web-based database collates app use statistics once reconnected to the internet.

Upon initial contact, researchers conducted a face-to-face, 20-minute session with participants to introduce and orient them to the app. After commencing one of the 2 character quests (ie, Ramone or Emily) and reviewing progress made, participants were then encouraged to take the app away and complete that character quest as well as their own quest (ie, input their own supportive people, strengths, worries, and goals) prior to the next appointment at 4 weeks. Participants received a standardized supportive text message weekly throughout the 4-week intervention period. These texts were to provide low-key well-being support, encourage app use, and troubleshoot any other issues with app access. For example, “Hi [name], [researcher’s name] here just wondering how you are going. Can you give us a call or text to arrange a time for a quick check-in? Also, let us know if you have any issues or questions about the app. Looking forward to hearing from you.” Texts were accompanied by a comic-style picture of the research team, which matched the graphic style of the app.

Participants received a phone credit voucher as reimbursement for their participation and to enable ongoing contact with the research team. Research officer contact guidelines for intervention and follow-up were strictly scripted in recognition that the human-support component of digital health approaches requires definition and concurrent evaluation [23]. Following completion of all study activities, participants were then gifted the mobile device if they did not already own one (they were not advised upon entry to the study that the devices would be given to them at completion).
Outcomes

Outcome measures were completed at baseline and 4 weeks. The primary outcome, psychological distress, was measured by the K10 [30]. The K10 is a 10-item measure of psychological distress that is sensitive to symptoms of both anxiety and depression [30]. Responses are on a 5-point Likert scale. K10 is one of the Australian Mental Health routine outcome measures and has been used in full and abbreviated forms (eg, K5, K6) in state- and nation-wide First Nations surveys [26,27,31,32]. For the period July 2012-June 2013, the Australian Mental Health Outcomes and Classification Network (AMHOCN) reports mean K10 scores for ambulatory patients with mood disorders (ie, outpatients returning to community after being treated acutely) across Australia of 27.6 (SD 8.5) upon return to community, 22.0 (SD 8.5) at a 91-day review, and 18.4 (7.6) upon discharge from outpatient service [33]. Considering these findings and those of our previous study with an Australian...
First Nations sample [18], a change or difference in K10 scores of 5 points would be considered clinically significant. Secondary outcome measures included the 2-question Patient Health Questionnaire (PHQ-2) with wording adapted by Brown and colleagues for First Nations Australians [34,35], the Generalized Anxiety Disorder short form (GAD-2) [35], the Leeds Dependence Questionnaire (LDQ) [36], the short form of the Alcohol Use Disorders Identification Test (AUDIT-C), the Drug Use Disorders Identification Test (DUDIT), the parent-rated Strengths and Difficulties Questionnaire (SDQ) [37], the General Help Seeking Questionnaire (GHSQ) [38], and a question adapted from the Client Service Receipt Inventory (CSRI) to determine the degree of concurrent mental health service use [39].

The PHQ-2 is a short-form version of the PHQ-9, which is designed to establish a psychiatric diagnosis of depression and has shown diagnostic, criterion, and construct validity [40,41]. The PHQ-9 has been tested in Indigenous groups and adapted to include simplified response categories [34,42,43], as well as specifically adapted for the central Australian context [34]. The 2-item PHQ-2 asks respondents to estimate the frequency of 2 symptoms (low mood and anhedonia) over the past 2 weeks on a 4-point Likert scale (0-3), with increasing scores indicating greater symptom severity. Total scores range from 0 to 6. A score of 3 or above has been shown to have a sensitivity of 83% and a specificity of 90% for the detection of major depressive disorder (MDD) [40]. The tool has also demonstrated utility for adolescent samples [44]. The GAD-2 asks respondents to estimate the frequency of 2 symptoms (nervousness and the ability to control worrying) over the past 2 weeks, with the 4 options and total scores ranging from 0 to 6 [45]. A cutoff score of 3 has been shown to have a sensitivity of 0.76 and a specificity of 0.81 [45]. The brief versions of the tools (ie, the PHQ-2 and the GAD-2) have previously been demonstrated to be reliable and valid for assessing change over time in clinical samples, with test-retest reliability of 0.79 for the PHQ-2 and 0.81 for the GAD-2 [35].

Three substance misuse screening tools were trialed. The LDQ was identified as a relatively brief (10-item) self-report measure sensitive to mild, moderate, and severe levels of dependence on alcohol and other drugs. Scores range from 0 to 30 and are intended to capture the graded intensity of the dependence syndrome. The ability to capture dependence severity simultaneously across all substance classes, including alcohol, sensitivity to change, high internal consistency (α=0.93), and clinical and research utility for young adults were identified as advantages [36,46]. The AUDIT-C is a shortened version of the AUDIT [47], which is the best practice tool currently recommended for alcohol screening in the general population and performs well for adolescents [48,49] and across ethnic groups [50]. The AUDIT-C consists of 3 items to determine the risk of hazardous and harmful drinking and alcohol dependence [47]. The AUDIT-C is scored on a scale of 0-12 points (scores of 0 reflect no alcohol use in the past year), and a cutoff score of 3 or more has been proposed when used with adolescents. As the AUDIT-C only examines alcohol use, the DUDIT was also trialed as a similar, 11-item tool for identifying other drug use patterns and drug-related problems. The first 9 items are scored on a 5-point scale ranging from 0 to 4, and the last 2 are scored on a 3-point scale with values of 0, 2, and 4. Total scores range from 0 to 44, with higher scores suggesting a more severe drug problem. The cutoff score for any type of problematic use (ie, harmful use, substance abuse, or dependency) is generally recommended as 6 for men and 2 for women. The DUDIT is reported to have adequate internal consistency (α=0.9+) and test-retest reliability (ICC=0.91; Pearson r=0.77), with favorable sensitivity and specificity reported in a review of its psychometric properties [51].

The parent-rated SDQ was trialed as an externally rated measure to minimize participant burden. The SDQ has been used with First Nations children in national Australian surveys and consists of 25 items measuring 5 areas of psychological adjustment: emotional symptoms, conduct problems, hyperactivity, peer problems, and prosocial behavior [52,53]. Evaluation of the psychometric properties for First Nations children recommended focusing on the total difficulties score and minimizing reliance on the peer relationships subscale [53].

Demographic information was collected at baseline. App feasibility and acceptability were assessed through app use data (including number of app opens, minutes spent in the app, page visits, help accessed, etc) and qualitative interviews at 4 weeks, exploring subjective experience. App ratings (ie, from 1=“didn’t like it at all” to 5=“liked it a lot”), look and style, content, overall app rating, supportive text message check-ins, and involvement in the study were also collected to assess acceptability. Study feasibility was assessed through recruitment and retention rates. Exit interviews investigated participants’ subjective experiences, likes and dislikes in terms of content, feel, look, and style, barriers and facilitators to use, suggested improvements, and engagement with the app. They also explored participants’ experiences of being involved in the study. Interview transcripts were analyzed by 4 authors (KD, JP, MS, and JF) using a deductive approach consistent with the above categories in line with research questions [54].

Field notes were examined for further indicators of acceptability, feasibility, any adverse events, and any feedback on study measures used to aid the interpretation of the data.

Feasibility was explored through the average number of minutes used and number of app opens over the 4-week period, as well as how many had completed their own quest. Acceptability was determined through thematic analysis of subjective experiences and high app ratings (ie, mean ≥3) in each of the areas measured.

Sample Size
For a paired-samples 2-tailed t test using the primary outcome K10, with a mean difference of 5 points and an SD of differences of 7, an α level of .05, and a power of 0.8, the total sample size required is 18. We anticipated a difference of at least 5 points would be indicative of clinically significant change [14,55].

Data Analysis
Demographic information, subjective ratings, and app use data are summarized using descriptive statistics (eg, means and SDs). Paired sample t tests were conducted for those with complete data (ie, baseline and follow-up assessment) for each of the
outcome measures, and within-group effect sizes (Cohen $d$) are reported. These were completer analyses, and there was no imputation of missing data. Exit interviews were deidentified and transcribed by an external transcribing service, then analyzed using a deductive approach to examine themes relevant to app use, likes, dislikes, facilitators and barriers to use, and suggested changes, as well as reflections on the study process. Four research team members coded the data separately and then discussed it to reach a consensus on relevant themes.

Results

Feasibility of Study and Outcomes

The study recruitment rate was 62.5% for those referred (35/56) and 81% for those able to be contacted (35/43) with a retention rate of 90% (30/33) for those who actually entered the study after screening.

All mental health indicators (ie, K10, PHQ-2, and GAD-2) showed a trend toward improvement on the repeated measures $t$ tests, with PHQ2 and K10 reaching clinical and statistical significance (see Table 1 and Figure 3). There was little reported substance use in this group, and no significant changes were observed for the AUDIT-C, DUDIT, or LDQ (see Table 1). The SDQ was only completed at both time points by 4 parents or guardians, and no significant differences were detected on this measure. There was little service use among the participants (see Table 2), but the most reported service used was psychologist or school counselor ($n=7$; 23% at baseline; $n=4$; 13.3% at follow-up) and Anglicare NT headspace ($n=5$; 16.7% at both baseline and follow-up).

In order to assess the feasibility of using a short form of the K10 (ie, the K5), the $t$ test was repeated using the K5 as the dependent variable. The change remained significant ($t_{29}=5.07$; $P<.001$). As shown in Table 1, mean scores at baseline were 11.97 and 9.07 at follow-up, with a large effect size (0.93).

Table 1. Means and $t$ tests for mental health and substance use measures at baseline and 4 weeks.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Baseline, mean (SD)</th>
<th>4 weeks, mean (SD)</th>
<th>Mean (SD) difference</th>
<th>$t$ test (df)</th>
<th>$P$ value</th>
<th>Effect size</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kessler Psychological Distress Scale (K10)</td>
<td>23.20 (7.39)</td>
<td>18.20 (5.20)</td>
<td>5.00 (5.92)</td>
<td>4.63 (29)</td>
<td>&lt;.001</td>
<td>0.85</td>
<td>2.79 to 7.21</td>
</tr>
<tr>
<td>Patient Health Questionnaire (PHQ-2)</td>
<td>2.53 (1.55)</td>
<td>1.70 (1.29)</td>
<td>0.83 (1.11)</td>
<td>4.09 (29)</td>
<td>&lt;.001</td>
<td>0.75</td>
<td>0.42 to 1.25</td>
</tr>
<tr>
<td>Generalized Anxiety Disorder short form (GAD-2)</td>
<td>1.53 (1.33)</td>
<td>1.2 (1.16)</td>
<td>0.33 (1.16)</td>
<td>1.58 (29)</td>
<td>.13</td>
<td>0.28</td>
<td>−0.10 to 0.77</td>
</tr>
<tr>
<td>Alcohol Use Disorders Identification Test (AUDIT-C)</td>
<td>0.44 (1.25)</td>
<td>0.26 (0.81)</td>
<td>0.19 (0.62)</td>
<td>1.55 (26)</td>
<td>.13</td>
<td>0.31</td>
<td>−0.06 to 0.43</td>
</tr>
<tr>
<td>Drug Use Disorders Test (DUDIT)</td>
<td>1.25 (2.78)</td>
<td>0.85 (2.77)</td>
<td>0.41 (1.80)</td>
<td>1.17 (26)</td>
<td>.25</td>
<td>0.23</td>
<td>−0.31 to 1.12</td>
</tr>
<tr>
<td>Leeds Dependence Questionnaire</td>
<td>0.57 (1.73)</td>
<td>0.57 (2.10)</td>
<td>0.00</td>
<td>0.0 (29)</td>
<td>$P&gt;.99$</td>
<td>0</td>
<td>−0.29 to 0.29</td>
</tr>
<tr>
<td>Strengths and Difficulties Questionnaire (SDQ-parent rated)–total difficulties ($n=4$)</td>
<td>37.00 (4.16)</td>
<td>32.75 (3.10)</td>
<td>4.25 (3.95)</td>
<td>2.15 (3)</td>
<td>.12</td>
<td>1.07</td>
<td>−2.03 to 10.53</td>
</tr>
</tbody>
</table>
Figure 3. Change over time for K10, PHQ-2, GAD-2, and Substance Use Measures (AUDIT-C, DUDIT, LDQ). AUDIT-C: Alcohol Use Disorders Identification Test-Concise; DUDIT: Drug Use Disorder Identification Test; GAD-2: Generalized Anxiety Disorder; K10: Kessler Psychological Distress Scale (10-item); LDQ: Leeds Dependence Questionnaire; PHQ-2: Patient Health Questionnaire.

Table 2. Service use at baseline and follow-up.

<table>
<thead>
<tr>
<th>Service used</th>
<th>Baseline participants, n</th>
<th>Follow-up participants, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hospital emergency department</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Hospital mental health ward</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Child and youth mental health service</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>GP or Aboriginal Community Controlled Health Service</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Psychologist or school counselor</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>Headspace</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>National helpline</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Other, for example, residential rehab, Clontarf</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

\(^a\)GP: general practitioner.

App Use

Descriptive statistics reflecting engagement with the app are summarized in Table 3 for 29 of the 30 participants (data missing for one participant and data not updated at follow-up for 5 participants, as phones were not available to connect to the internet, so that is likely to be a conservative estimate for 5 participants). Participants spent on average 37 minutes in the app. Three participants (10%) only opened the app once (presumably at baseline with the researcher). The app help page was accessed by 14 (47%) young people at least once, of whom one accessed the Australian Crisis Lines webpage and another accessed the Kids Helpline phone number from the app. All but 3 participants accessed their own quest to input their own story and goals, and 16 (55%) completed all 10 levels. For the levels, the most amount of time on average was spent in the fishing game on “Ramones Quest” (mean 1.30, SD 2.85) and the bubbles game on “My Quest” (mean 1.28, SD 5.98).

Separate regression analyses were conducted to see if age, time spent in the app, and number of app opens predicted mental health outcomes (K10, GAD-2, or PHQ-2). Age, use time, and number of app opens did not predict scores on any of the 3 mental health outcomes (K10, GAD-2, or PHQ-2). Simple bivariate correlations also showed no significant relationship between age and the number of app opens (r=0.01; P=.95) or number of minutes spent in the app (r=0.04; P=.85).
Table 3. App use statistics for sample (n=29).

<table>
<thead>
<tr>
<th>Measure</th>
<th>Mean (SD)</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minutes spent in app</td>
<td>37.87 (27.67)</td>
<td>10.33</td>
<td>143.28</td>
</tr>
<tr>
<td>Number of app opens</td>
<td>5.69 (3.78)</td>
<td>1</td>
<td>16</td>
</tr>
<tr>
<td>Number of times help accessed</td>
<td>0.90 (1.21)</td>
<td>0</td>
<td>5</td>
</tr>
</tbody>
</table>

User Experience Ratings

Users generally rated the app positively, with mean ratings around 4 (on a scale from 1-5) for each of the 5 items: look and style, content, overall rating, check-in text messages, and study acceptability (see Table 4).

Table 4. Mean acceptability ratings for elements of the app and study
cia.

<table>
<thead>
<tr>
<th>Item</th>
<th>Mean (SD)</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Frequency scored 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1: Look and style</td>
<td>4.28 (0.70)</td>
<td>3</td>
<td>5</td>
<td>12</td>
</tr>
<tr>
<td>Q2: Content</td>
<td>4.38 (0.86)</td>
<td>2</td>
<td>5</td>
<td>17</td>
</tr>
<tr>
<td>Q3: Overall rating</td>
<td>4.59 (0.50)</td>
<td>4</td>
<td>5</td>
<td>17</td>
</tr>
<tr>
<td>Q4: Check-ins</td>
<td>4.12 (0.83)</td>
<td>3</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>Q5: Study acceptability</td>
<td>4.62 (0.62)</td>
<td>3</td>
<td>5</td>
<td>20</td>
</tr>
</tbody>
</table>

cia Ratings on scale of 1=“didn’t like it at all” to 5=“liked it a lot.”

User Feedback

In terms of subjective experience, participants reported that it helped them feel relaxed, happy, calm, or safe (P23, P35, P30). It helped one reflect on their life or “check back in with yourself” (P27). Another said it helped them “feel better about myself” (P29). Participants suggested the app could help them or their mates get help in tough times. “I used it to like give me strategies and that to deal with tough times” (P35). “It was just a good experience. You need to focus on yourself sometimes and make goals for yourself. And, yeah. Especially if you’re going through a rough time and all that, and you forget people are there for you. The app says, ‘Do you have people there for you?’ And you remember. Like, ‘Oh, yeah. There is people still there for you’” (P27).

Overwhelmingly, playing the games (P5, 9, 22, 23, 24, 25, 26, 29, 31, 32, 33, 35, 38) was mentioned as a favorite part of the app, as was watching the videos (P9, 11, 22, 24, 26, 27, 29, 38). At least five participants liked remembering and identifying their strengths (P24, 26, 29, 38, 33). The rewards (P27, 29, 38), character stories (P1, P22, P23, P29), adding photos of family (P26) and people who care about you (P5, 35, 24, 27), and listening to the music (P26) were also mentioned as being liked.

Participants liked the look and feel of the app, it “felt natural. It wasn’t something that was like foreign” (P38). Others said it was “comfortable and homey” (P24) and “traditional” (P30). One participant suggested it “reminded me of home” (P24). Participants liked how colorful it was and the use of cultural imagery and stories “The colors are really good...and really effective. I mean the Indigenous side of it and all that” (P27). “It was nice having them talk about their culture and stuff” (P1). “I like all the little pictures” (P23).

Reasons for use included that it helped them feel relaxed, helped if they were bored, used it if their other phone was flat, checked in, and helped if they were feeling down. “I really like it because it’s helped me like get happy. Like if I was sad, or lazy, or something, I’d just look over the app and it would just get me up and going.” (P8). The text messages sent as reminders were thought to prompt use. Participants generally thought the app would be useful for First Nations young people aged 10-16 or 18 years while one or two thought it might also be useful for adults or younger kids and could even be “relatable for other people of different heritage” (P38).

Barriers to use included not having the app on their own phone (due to it only being available on Android devices for the trial), forgetting about it, its repetitiveness, and having better things to do or no time. A couple of participants reported getting lost or confused, but most reported that the app was easy to use and understand. Others suggested that not knowing what else to do once they had finished all the quests prevented them from using the app again. “So, when you completed both stories, I didn’t really know what to do after that. I kind of got stuck” (P8). While some enjoyed the games, others said they were too easy or too slow and preferred more challenging games to be included. A few participants reported that they did not really read the stories; they just skipped over them. Some did not see parts of the app, particularly the “get help” page. Some admitted that they had not used the app much or hadn’t used it again once they completed it. When they did use it, it was commonly only once or twice a week and often at night.

Suggestions were made to help overcome some of these reported barriers, which overwhelmingly included adding more content and variety, including additional characters or stories and more interactive or challenging games. The stories could include both older and younger characters, role models (eg, famous people and their stories), and more in-depth, place-specific stories. “Yeah, ‘cause it went really quick, them 2 journeys. I kept watching them and watching them over again” (P23). Other suggestions were to be able to “share your own story” (P8) or...
“do it for your mum or your dad as well” (P24). One participant suggested having a bit more instruction or a tutorial on how to work through the app. Additional audio was mentioned by a few participants that could be used in various ways. For example, a voice with encouraging feedback (eg, “Well done. You’re doing great” [P24]) or to read out the stories for those who have trouble or do not like reading (P35, 31). Notifications from the app were generally suggested as useful reminders to use or revisit the app, particularly after 2 weeks or when new content becomes available. One person also suggested some more customization settings would promote use, and another wanted more real-life videos (with movement rather than static imagery).

Participants often felt a sense of pride at having been involved in the project “I actually felt really good for this. Like, one of the first people to try.... Like, Holy Moley! I’m one of the first people to use the app” (P32). Most participants in the school or educational setting were happy to be interrupted in class to participate and reported that the process was good and that they felt they could say no to participating.

In summary, these results suggest that the app was thought to be acceptable, (“It’s a good app” [P1]; “I enjoyed it” [P38]; “I really like it” [P8]), easy to use (“I liked how it was really simple to do” [P35]), and culturally relevant (“it gives out like a cultural–culture pictures” [P23]; “it was nice having them talk about their culture” [P1]; “It made me feel good like that there was an app for us” [P24]) “And really effective. I mean the Indigenous side of it and all that” [P27]) and useful (“it made me feel more–felt better about myself” [P29]; “It kind of like calmed me down.... I felt like, better and sort of safer” [P30]). While it seemed that the app was not overly engaging, with some only using it once or twice, improvements were suggested that might increase engagement, including app reminders, more variety in the content, the addition of audio, more interactivity, and more stories, videos, and challenging games.

Discussion

Principal Results

This study assessed the feasibility, and acceptability of the newly developed AIMhi-Y app. Results demonstrated that the AIMhi-Y app is a feasible and acceptable approach to improving mental health for First Nations youth. Findings showed improved mental health outcomes for participants following 4 weeks of AIMhi-Y app use. Statistically and clinically significant improvements in psychological distress and depressive symptoms were demonstrated over the study period. However, due to the lack of appropriate control, the role the app played in these findings is unclear. These results are encouraging, as no adverse events were forthcoming and the app appeared to provide benefits. The app was also deemed acceptable by study participants in their ratings and reviews of the app. Young people reported that they found the app easy to use, culturally relevant, and useful. Engagement with the app was restricted to around 37 minutes on average, with an average of 6 opens during the 4 weeks. Suggestions were made for increasing engagement with the app that included notifications, an increased and greater variety of content, including more challenging games, and additional videos. The feasibility of the study was also demonstrated with a 62% recruitment and 90% retention rate observed and high study acceptability ratings.

Comparison With Prior Work

The findings of significant reductions in psychological distress and depressive symptoms mirror similar recent findings in a trial of a suicide prevention app for Aboriginal and Torres Strait Islander young people, the iBobby app [16]. iBobby demonstrated significant improvements after 6 weeks, and our study showed significant improvements after 4 weeks. The study also demonstrated no significant deterioration in the specific domains measured. Taken together, these results provide encouraging support for the utility of these dMH apps and suggest that appropriately designed dMH apps such as these are a feasible and acceptable means of lowering symptoms for mental health disorders in regional and remote First Nations communities.

The findings also suggest that the measures used were appropriate and feasible for detecting change over time in this sample. They were well received, as demonstrated by the young people’s willingness to complete and confidence in responding, as well as their ability to demonstrate statistically significant improvements over a relatively short period. The brief versions of the tools have previously been demonstrated to be reliable and valid for assessing change over time in clinical samples [35], and this appears to hold true for our adolescent sample. Given this finding, the use of the brief versions is recommended to limit participant burden in future studies. The utility of the drug and alcohol use measures, however, was more difficult to determine in our sample of young people who demonstrated minimal substance use generally. The degree to which these measures can demonstrate change in this population over this period is, therefore, less clear. Previous research has demonstrated that changes in AUDIT-C scores over one year do reflect changes in drinking for adults and can predict future problematic drinking in adolescents [48,56]. Thus, the use of a brief assessment such as the AUDIT-C is likely to be adequate for this group. However, the SDQ may not be a feasible measure for future trials. It provided limited outcome data and was completed at both time points by only 4 parents or caregivers. Significant investment may be needed to ensure follow-up, especially for those young people living remotely or somewhat independently from their primary caregiver or parent.

Despite the positive results in well-being measures, this study also reflects previous literature suggesting that there is often high enthusiasm for dMH but low uptake and sustained use of mobile mental health apps [20]. While ratings and subjective reports indicated that the app was acceptable to participating young people, use data and subjective reports reflected modest user engagement. Time spent in the app averaged 37 minutes, and there were an average of 6 app opens per user over the 4-week period. Use was comparable to a previous study of another First Nations–specific app targeting suicide prevention (ie, iBobby), in which participants used the app on average for 73 minutes with 12 logins on average over 6 weeks [15]. AIMhi Stay Strong MCP is a brief intervention that is the foundation for the AIMhi-Y app [18]. AIMhi Stay Strong MCP was
developed to be administered over two 20-minute sessions. A recent RCT trialing AIMhi Stay Strong MCP in the form of the AIMhi Stay Strong app demonstrated significant improvements in distress and depression after 2 clinician-supported 20-minute sessions with the app [14]. While most of the app use in this study was self-directed, it is feasible that 40 minutes of use may be sufficient to elicit some change. It is also important to note that our intervention sent only one SMS message each week and that more frequent messages may have increased engagement.

In the context of other apps, this level of engagement is not unusual. Industry market research has reported that only 38% of users engage with a particular app more than 11 times, and 24% of users abandon an app after only one use [57]. Several ways of increasing user engagement have been suggested, which can include both improvements to the features of the app and the systems around them [29]. Improving app features might focus on increasing appeal, improving usability, or enhancing adherence [29]. Things like having a simple and intuitive user interface, delivering concise information in app messages, personalization and customization, incentives, updates, and new content, and good onboarding can improve user retention [57]. Others suggest offering a challenge may improve engagement over short periods, thereby allowing some therapeutic benefit and the potential for developing it into a habit [58]. In line with the above findings, participants in this study suggested including notifications, additional and greater variety of content, more challenging games, audio, and additional videos as potential ways to increase engagement.

Gamification (ie, the use of game design elements and features in nongame contexts) is commonly offered as a key opportunity to improve both appeal and adherence, which fits with the preferences of participants in this study, who favored gamification elements such as rewards, mini-games, and character stories [29,59]. Gamification features (eg, a narrative or theme, progress feedback, rewards, leaderboards, badges, customizable avatars, personalization, social interaction, competition, or user choice) need to be well designed, integrate seamlessly with the technology, and have a clear purpose and user involvement early in their design for greatest feasibility and acceptability [60,61].

Low engagement may be one of the reasons for equivocal reports when it comes to app effectiveness. If a user fails to engage with an app, it is unlikely to lead to benefits or improved outcomes. Issues with intervention engagement are not restricted to digital interventions, and similar challenges can occur with face-to-face delivery [58]. Nevertheless, the benefits of mobile apps are more consistently achieved in the context of human support [23]. Recent evidence suggests that the nature of that support is an important element in improving outcomes. Support that targets engagement alone is not sufficient to improve clinical outcomes beyond those of unsupported interventions [8]. On the other hand, support that addresses the reasons people might fail to benefit from dMH interventions, such as deficiencies in usability, engagement, fit, knowledge, or implementation, is likely to be more effective. A focus on both increasing engagement and furthering knowledge and implementation of the skills learned in the dMH intervention is needed [8]. Providing clear guidelines for how, when, and why to use support might also be of benefit [23] and highlight new research questions to be answered beyond this study.

Further co-design of the human support and implementation components of this intervention is likely to strengthen its effectiveness and is currently underway, including contextualizing it to other Australian locations given there is regional variation in First Nations cultures across Australia. The use of user-centered design that emphasizes deep engagement with key stakeholders and their organizational and social contexts has been recommended [62] and will be applied to this intervention. This is particularly important as research suggests that optimization of the design of human support services may have a greater impact on clinical outcomes than does the design of the technologies [62]. For providers (care managers, physicians, and mental health providers), the dMH tool must fit into their workflows and offer some meaningful benefit rather than just adding another task to their workdays in order to aid implementation [23]. While optimizing the human support component and considering ease of implementation are important, there is also a need to further develop the appeal of dMH interventions such as this to increase reach to the 80% of young people who are not already accessing professional treatment [63].

Limitations
While the findings from this study appear promising, they must be considered in light of several limitations. As mentioned, the lack of a control group limits the degree to which we can attribute the outcome changes to the intervention. Future research using a hybrid trial design is recommended to investigate effectiveness and implementation success concurrently. Such a design might address the increasing evidence of a research-to-practice gap in this field of technology-enabled services (or dMH services) [23]. In addition, the study sample was relatively homogenous in that it was derived primarily from First Nations young people engaged in school and residing in an urban setting (Darwin). While we attempted to recruit a diverse sample through the inclusion of participants attending a mental health service and a residential drug rehabilitation service (from remote communities), the majority of participants recruited were through the school site, thus limiting the generalizability of the findings. On the other hand, if a greater diversity of participants from additional regions is included in future studies, it may require tailoring to context through the co-design of additional app elements or content relevant to the First Nations cultures in that region. Participants were also willing volunteers who were chosen for referral by participating key site contacts, and thus the resulting sample may have resulted in selection bias.

Conclusions
This study showed positive effects on well-being following 4 weeks of AIMhi-Y app use. While the lack of a control group tempered the strength of these findings, the relative feasibility and acceptability of the intervention were demonstrated. High approval ratings were observed alongside modest engagement, suggesting strategies for increasing engagement, such as well-designed human support and technological and content
improvements, may lead to increased use and thus increased benefit. This study supports earlier research suggesting that dMH apps that are appropriately designed with and for the target populations are a feasible and acceptable means of lowering symptoms for mental health disorders among First Nations young people.

**Acknowledgments**

We would like to acknowledge all the young people, parents, teachers, support people, and members of our ERG who made this research possible. The authors would like to thank Senior Indigenous Researcher Patj Patj Janama Robert Mills for supporting this research. Funding was provided by Suicide Prevention Australia Limited through an Innovation Research Grant.

**Conflicts of Interest**

None declared.

**References**


15. Tighe J, Shand F, McKay K, Mcalister TJ, Mackinnon A, Christensen H. Usage and acceptability of the iBobbly app: pilot trial for suicide prevention in Aboriginal and Torres Strait Islander youth. JMIR Ment Health 2020;7(12):e14296 [FREE Full text] [doi: 10.2196/14296] [Medline: 33258782]


Abbreviations

AIMhi: Aboriginal and Islander Mental Health Initiative
AIMhi-Y: Aboriginal and Islander Mental Health Initiative for Youth
AUDIT-C: Alcohol Use Disorders Identification Test (short form)
CBT: cognitive behavioral therapy
CSRI: Client Service Receipt Inventory
dMH: digital mental health
DUDIT: Drug Use Disorders Identification Test
ERG: Expert Reference Group
GAD-2: Generalized Anxiety Disorder Scale (2-item)
GHSQ: General Help Seeking Questionnaire
IYRG: Indigenous Youth Reference Group
K10: Kessler Psychological Distress Scale (10-item)
LDQ: Leeds Dependence Questionnaire
MCP: motivational care planning
NT: Northern Territory
PHQ-2: Patient Health Questionnaire (2-item)
RCT: randomized controlled trial
SDQ: Strengths and Difficulties Questionnaire

©Kylie M Dingwall, Josie Povey, Michelle Sweet, Jaylene Friel, Fiona Shand, Nickolai Titov, Julia Wormer, Tamoor Mirza, Tricia Nagel. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 07.06.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
A Health App for Evidence-Based Postpartum Information: Development and Validation Study

Máyra Cármen Silva de Medeiros1, MSc; Maiara Costa de Oliveira2, MSc; Leonandro Valério Barbosa Gurgel3, BSc; Anna Giselle Câmara Dantas Ribeiro Rodrigues3, PhD; Maria Thereza Albuquerque Barbosa Cabral Micussi2, PhD; Adriana Gomes Magalhães1, PhD

1Faculty of Health Sciences, Universidade Federal do Rio Grande do Norte, Santa Cruz, RN, Brazil
2Department of Physiotherapy, Universidade Federal do Rio Grande do Norte, Natal, Brazil
3Department of Information Technology, Universidade Federal do Rio Grande do Norte, Natal, Brazil

Corresponding Author:
Máyra Cármen Silva de Medeiros, MSc
Faculty of Health Sciences
Universidade Federal do Rio Grande do Norte
R. Passos de Miranda
Santa Cruz, RN, 59200-000
Brazil
Phone: 55 (84) 3291 2411
Email: mayracmed@gmail.com

Abstract

Background: After childbirth, women undergo substantial physical and emotional changes. Therefore, it is important to provide them with information that helps them identify what is expected during this stage, as well as signs and symptoms that indicate complications after they have been discharged from the hospital.

Objective: This study aimed to develop a health app—Towards Motherhood—that provides evidence-based information about the postpartum period and evaluate the usability of the app with the target population.

Methods: This was a validation study involving 80 participants, including 24 professionals from the obstetric health field, 15 professionals from the technology field, and 41 postpartum women. The app was developed using React Native technology. Health professionals evaluated the app’s content using the Content Validity Index, technology professionals completed a validated evaluation to assess the appearance of the app, and postpartum women completed the System Usability Scale (SUS) to measure the usability of the app.

Results: The measurement of content validity using a Likert scale obtained an approval score of 99%. Regarding the app’s appearance, 92% of responses were positive, reflecting favorable approval. The SUS usability score was 86.2, which represents excellent acceptance.

Conclusions: The Towards Motherhood mobile app is a valid tool for promoting self-care during the postpartum period. The app’s evidence-based information, user-friendly design, and high usability make it an essential resource for women during this critical stage of their life.

Introduction

In the first few days after childbirth, women undergo substantial physical and emotional changes, making it important to provide them with information to identify what to expect during this stage and the signs and symptoms of possible complications after hospital discharge, such as bleeding, pain, and urinary tract infection. Accessible and reliable information, prevention, and care for complications are essential and must be adopted [1,2]. The development of information and communication technology has provided new ways to improve users’ quality of life by monitoring their health status [3]. Telecare, telehealth, and...
mobile health (mHealth) are components of innovative and frequently used methods [4-5]. These tools can potentially improve people’s health status, are considered tools of great utility for solving or reducing the health problems of individuals or populations, and are user-friendly [6].

It is evident that today, there has been a growth of mobile technologies and apps (mHealth) that contribute to the production of a new modality of health care, in which information regarding people’s health is relevant and universal [7,8].

Given the above, this work aimed to develop a health app—Towards Motherhood—with information about the puerperal phase. Its content was elaborated based on updated scientific knowledge to provide safe content on topics pertinent to this phase. This was then followed by an evaluation on the usability of the app with the target population.

Methods

Characterization of Research

This was a validation study conducted at the Januário Cicco Maternity School in the city of Natal, Rio Grande do Norte, Brazil.

Ethics Approval and Informed Consent

This study was submitted to and approved by the Research Ethics Committee of the University Hospital Onofre Lopes of the Federal University of Rio Grande do Norte (CAAE: 38145320.2.0000.5537). All research participants voluntarily agreed to participate and signed the free and informed consent form.

Population and Sample

To develop the app, a research group consisting of 2 expert physiotherapists, 3 undergraduate students (one each from IT, physiotherapy, and graphic design), 2 physiotherapy professors, and 1 IT professor held weekly meetings throughout 2021. The app was constructed through database research; group discussions for knowledge translation; evidence-based content preparation and review of screen prototypes; implementation; and subsequent validation of content, functionality, and design.

The evaluation inclusion criteria for health professionals required experience in the obstetric health area and at least a specialist title, whereas the inclusion criteria for IT professionals required previous experience developing mHealth, React Native, or front-end software based on a previous study of mobile app validation. Postpartum women who had given birth at the Januário Cicco Maternity School and were still hospitalized in the institution were included, whereas the exclusion criteria were not agreeing to participate in the research or not signing the free and informed consent form.

Sample Size

A total of 80 participants were included: 39 professionals (24 health professionals with specialization in obstetrics and 15 IT professionals) and 41 postpartum women. The number of research participants was determined based on articles that use mHealth technology. It has been emphasized that there is no consensus in the international literature on the minimum number of judges, but there is agreement on the importance of clinical experience in the formation of a profile of expertise, as well as the need to balance clinical experience and solid academic training [9-12]. The participants were selected by convenience sampling according to previous research [13].

Development

The app was developed using React Native, a JavaScript library created by Facebook (Meta Platforms Inc) to build mobile apps for the Android operating system. The development process consisted of 4 main stages, as illustrated in Figure 1.

Figure 1. The app development process.
Validation
The expert professionals were contacted via email or telephone and invited to participate in the study. They were provided with information about the research, and upon agreement, they signed the free and informed consent form. The professionals were given access to the app content, evaluated it, and provided feedback.

Postpartum women were approached in the ward, and after signing the free and informed consent form, they were given access to the app through a tablet provided by the evaluator for 20 minutes. After use, participants were requested to complete a form on the Google Forms platform using the tablet.

The process was divided into three stages: (1) evaluation of the content by health care professionals, (2) evaluation of the app’s appearance by IT professionals, and (3) usability assessment with postpartum women.

Content Assessment
The content of the app was obtained through research in the main databases of Scientific Electronic Library Online, ScienceDirect, Cochrane, Web of Science, Scopus, and MEDLINE. The following Descriptors in Health Sciences and Medical Subject Headings were used: “Saúde da Mulher” (“Women’s Health” in Portuguese) and “Período Pós-Parto” (“Postpartum Period” in Portuguese).

Documents with up-to-date scientific evidence were selected. The content to be inserted in the app was discussed among the members of the research group, most of whom had experience in the area of women’s health. Subsequently, the information was transcribed into language that was easy for the target audience to understand.

The health professionals evaluated the content within the app using a Likert scale with responses ranging from 1 (strongly disagree) to 5 (strongly agree). The evaluated content themes are shown in Figure 2. The Content Validity Index was used to measure agreement on the scores given by specialists for each item, with a final score given as a percentage, which should be greater than 78% [8]. The evaluators with expertise in obstetrics participated in this stage.

Appearance Assessment
The app’s appearance was evaluated through the User Experience Questionnaire with 6 questions: “Is the language used in the app easy to understand?” “Are the features used in the app implemented correctly?” “Are the features used in the app conducted comprehensively?” “Is the app interface attractive?” “Is the app easy to manage?” and “Does the app...
provide help in a non-tiring way?” The answers were measured using a Likert scale with 5 possible responses, as in the previous stage. IT professionals participated in this stage, and a minimum of 78% of positive responses was required for approval [6].

Usability Assessment

Postpartum women evaluated the usability of the application using the System Usability Scale (SUS), a questionnaire that has been translated and validated in Portuguese. It consists of 10 questions, with 5 positively worded statements and 5 negatively worded statements: “I think I would like to use this system frequently,” “I find the system unnecessarily complex,” “I found the system easy to use,” “I think I would need help from a person with technical knowledge to use the system,” “I think the various functions of the system are very well integrated,” “I think the system presents a lot of inconsistency,” “I imagine people will learn how to use this system quickly,” “I found the system clumsy to use,” “I felt confident using the system,” and “I had to learn a lot of new things before I could use the system.” The evaluators responded using a 5-point scale ranging from 1 (strongly disagree) to 5 (strongly agree). The overall score was then calculated on a scale from 0 to 100 points, with a cutoff point of 68 to consider the app as being usable [9,10]

Statistical Analysis

The sample data were analyzed using the SPSS statistical software (version 20.0; IBM Corp) with a significance level of 5%. The Content Validity Index and User Experience Questionnaire score were calculated by summing up the values of the answers and presenting them as a percentage. The SUS score was calculated as follows: for odd items, 1 was subtracted from the position on the scale, and for even items, 5 was subtracted from the position on the scale; then, all items were summed and multiplied by 2.5 to obtain the overall usability score of the system.

Results

For the development of the app, an integrative review was conducted in the main databases of Scientific Electronic Library Online, Science Direct, Cochrane, Web of Science, Scopus, and MEDLINE.

Initially, the app was named “Towards Motherhood” and was designed for offline use with free access on the Android platform. The app’s main menu offers 5 topics for exploration: Emotional Changes, Breastfeeding, Abdominal and Uterine Muscles, Varicose Veins, and Edema, as shown in Figure 2. Each topic includes subtopics for easy navigation and access to information.

The content validation process included 24 health professionals with expertise in obstetrics, of which 83% (n=20) were women. The group included 33% (n=8) physiotherapists, 29% (n=7) medical professionals, and 38% (n=9) nurses. More than half (n=13, 54%) had a specialization degree and experience in the public health system, 25% (n=8) had a master’s degree, 8% (n=2) had a doctorate degree, and the remaining 13% (n=3) were undergraduate students. Table 1 displays the answers and comments provided by the health professionals, with the suggestions discussed by the research group and accepted based on pertinence and scientific evidence. Breastfeeding was the topic with the highest number of suggestions, whereas the content on the postpartum period and its stages had a 100% agreement and no suggestions. Content validation was conducted through the Likert scale, obtaining a score of 97%.

The appearance of the app was evaluated by 15 IT professionals, comprising of 67% (n=10) male and 33% (n=5) female participants. They commented that the app was easy to use and had good understanding of the functionalities of the elements. Whenever there was a disagreement among the specialists on any item of the app, they proposed a new statement or new title for the menu, recommended the inclusion of additional information, or suggested the inclusion of a new item. The approval rate was 92% positive responses, which was favorable. The responses are presented in Table 2.
<table>
<thead>
<tr>
<th>Subject</th>
<th>Strongly disagree, n (%)</th>
<th>Partially disagree, n (%)</th>
<th>Neutral, n (%)</th>
<th>Partially agree, n (%)</th>
<th>Strongly agree, n (%)</th>
<th>Suggestions</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Puerperal blues</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>3 (13)</td>
<td>21 (87)</td>
<td>Emphasize that the time to rest and sleep is essential, and the support network is essential to take care of the baby in these moments</td>
<td>Accepted</td>
</tr>
<tr>
<td>Postpartum depression</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>8 (33)</td>
<td>16 (67)</td>
<td>Seek professional help</td>
<td>Accepted</td>
</tr>
<tr>
<td>Breastfeeding</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>7 (29)</td>
<td>17 (71)</td>
<td>Create a topic on how to make the correct handle</td>
<td>Accepted</td>
</tr>
<tr>
<td>Uterine involution</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>4 (17)</td>
<td>20 (83)</td>
<td>Seek medical attention if the pain is not ceasing</td>
<td>Accepted</td>
</tr>
<tr>
<td>Abdominal diastasis</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>6 (25)</td>
<td>18 (75)</td>
<td>Show images of some movements used to minimize postpartum diastasis</td>
<td>Evaluation required</td>
</tr>
<tr>
<td>Urinary incontinence</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>4 (17)</td>
<td>20 (83)</td>
<td>The term “postpartum incontinence of urine” looks like a classification</td>
<td>Accepted</td>
</tr>
<tr>
<td>Scars arising from childbirth</td>
<td>0 (0)</td>
<td>1 (4)</td>
<td>0 (0)</td>
<td>4 (17)</td>
<td>19 (79)</td>
<td>Hygiene practices</td>
<td>Accepted</td>
</tr>
<tr>
<td>Lochia</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>8 (33)</td>
<td>16 (67)</td>
<td>Details about duration and warning signs</td>
<td>Accepted</td>
</tr>
<tr>
<td>Intimate hygiene</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>2 (8)</td>
<td>22 (92)</td>
<td>Avoid tampons</td>
<td>Accepted</td>
</tr>
<tr>
<td>Care in defecation</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>2 (8)</td>
<td>22 (92)</td>
<td>Better explanation about the position of squats</td>
<td>Accepted</td>
</tr>
<tr>
<td>Varicose vein</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>3 (13)</td>
<td>12 (87)</td>
<td>Orientation about the importance of talking with the doctor about compression stockings</td>
<td>Accepted, with physiotherapists added to the suggestion</td>
</tr>
<tr>
<td>Edema</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>3 (13)</td>
<td>12 (87)</td>
<td>Orientation about water intake</td>
<td>Accepted</td>
</tr>
<tr>
<td>Healthy habits</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>4 (17)</td>
<td>20 (83)</td>
<td>Well-being and leisure</td>
<td>Accepted</td>
</tr>
<tr>
<td>Postpartum sexual activity</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (4)</td>
<td>3 (13)</td>
<td>20 (83)</td>
<td>Importance of talking about contraception</td>
<td>Accepted</td>
</tr>
<tr>
<td>Back pain in postpartum period</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (4)</td>
<td>0 (0)</td>
<td>23 (96)</td>
<td>Exercise videos</td>
<td>Evaluation and prescription of exercises should be performed individually</td>
</tr>
</tbody>
</table>
Table 2. Responses of IT professionals on the app’s appearance (n=15).

<table>
<thead>
<tr>
<th></th>
<th>Strongly disagree, n (%)</th>
<th>Partially disagree, n (%)</th>
<th>Neutral, n (%)</th>
<th>Partially agree, n (%)</th>
<th>Strongly agree, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is the language used in the app easy to understand?</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (7)</td>
<td>4 (27)</td>
<td>10 (67)</td>
</tr>
<tr>
<td>Are the features used in the app implemented correctly?</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>7 (47)</td>
<td>8 (53)</td>
</tr>
<tr>
<td>Are the features used in the app conducted comprehensively?</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>5 (33)</td>
<td>10 (67)</td>
</tr>
<tr>
<td>Is the app interface attractive?</td>
<td>2 (13)</td>
<td>3 (20)</td>
<td>5 (33)</td>
<td>5 (33)</td>
<td></td>
</tr>
<tr>
<td>Is the app easy to manage?</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (7%)</td>
<td>3 (20)</td>
<td>11 (73)</td>
</tr>
<tr>
<td>Does the app provide help in a non-tiring way?</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>5 (33)</td>
<td></td>
<td>10 (67)</td>
</tr>
</tbody>
</table>

The usability of the application was assessed by 41 postpartum women aged 18 to 40 years, with the majority (n=27, 66%) having completed high school education, followed by 19% (n=8) who were literate and 15% (n=6) with higher education. The majority (n=25, 61%) of the participants were single and the rest (n=16, 39%) were in a stable union. The users’ feedback is presented in Table 3, where their comments on the Google Forms questionnaire included “I found it very informative,” “Good and easy to use,” “I enjoyed the experience,” “The information was very useful,” and “Very good.” Usability was evaluated using the SUS, which yielded a score of 86.2, indicating excellent acceptance. However, the available version is a prototype developed for app validation testing and is not yet available for free access.

Table 3. User responses on the usability of the app (n=41).

<table>
<thead>
<tr>
<th></th>
<th>Strongly disagree, n (%)</th>
<th>Partially disagree, n (%)</th>
<th>Neutral, n (%)</th>
<th>Partially agree, n (%)</th>
<th>Strongly agree, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I think I’d like to use this system often</td>
<td>1 (2)</td>
<td>3 (7)</td>
<td>1 (2)</td>
<td>11 (27)</td>
<td>25 (61)</td>
</tr>
<tr>
<td>I find the system unnecessarily complex</td>
<td>30 (73)</td>
<td>3 (7)</td>
<td>3 (7)</td>
<td>3 (7)</td>
<td>2 (5)</td>
</tr>
<tr>
<td>I found the system easy to use</td>
<td>3 (7)</td>
<td>1 (2)</td>
<td>0 (0)</td>
<td>5 (12)</td>
<td>32 (78)</td>
</tr>
<tr>
<td>I think I would need help from a person with technical knowledge to use the system</td>
<td>24 (59)</td>
<td>3 (7)</td>
<td>0 (0)</td>
<td>10 (24)</td>
<td>4 (10)</td>
</tr>
<tr>
<td>The various functions of the system are very well integrated</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>5 (12)</td>
<td>36 (88)</td>
</tr>
<tr>
<td>The system presents a lot of inconsistency</td>
<td>34 (83)</td>
<td>3 (7)</td>
<td>1 (2)</td>
<td>1 (2)</td>
<td>2 (5)</td>
</tr>
<tr>
<td>I found the system clumsy to use</td>
<td>38 (93)</td>
<td>1 (2)</td>
<td>0 (0)</td>
<td>1 (2)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>I felt confident using the system</td>
<td>1 (2)</td>
<td>1 (2)</td>
<td>1 (2)</td>
<td>3 (7)</td>
<td>35 (85)</td>
</tr>
<tr>
<td>I had to learn several new things before I could use the system</td>
<td>23 (56)</td>
<td>8 (20)</td>
<td>0 (0)</td>
<td>5 (12)</td>
<td>5 (12)</td>
</tr>
</tbody>
</table>

**Discussion**

**Principal Findings**

Developing a technology to facilitate the acquisition of evidence-based content for a stage of life that brings countless doubts to women is of utmost relevance. This fact can be supported by the identification that such technologies are scarcely available in the main web stores and are not widely published in the major health journals [14].

The main objective of this study was not only to create an app about postpartum care but also to develop a technology that aligns with self-care for women, as empowering these women is essential to avoiding complications in the postpartum phase. As a refinement of this app, health professionals with extensive experience in obstetrics were able to give their opinion on the content, as well as find the necessary areas of improvement using their practical experience to determine what they perceive to be the main difficulties and doubts of puerperal women. Existing work in this area remains incomplete, with a limited sample size and a need of further investigation [15].

Knowledge translation is a means to communicate scientific evidence in an effortless way, with the objective of being effectively understood and applied in real life and influencing the creation of new products and technologies. In the context of education and health promotion, care should be taken...
regarding the adequacy of the language used to facilitate understanding. Popularly used words should preferably be used, and technical terms should be restricted to what is strictly necessary [12,16].

Learning is directly influenced by the social and cultural beliefs of the environment in which it is embedded, and therefore, the content of the app was selected with the aim of not disrespecting these issues. As a refinement, an attempt was made to adapt the guidance to all audiences without losing its scientific nature, which is supported by relevant literature. The content is presented not only in text but also in images to facilitate understanding.

According to literature, a SUS score above 68 indicates acceptable usability, whereas a score of 85 or above is related to excellent approval of software or applications. The mean SUS score for the Towards Motherhood app reached these parameters, as seen in a broad examination of the SUS [15,17].

Analysis of the SUS items showed greater variance in responses for “I think I would need help from a person with technical knowledge to use the system” and “I felt confident when using the system.” This result could be attributed to the low level of education of users, highlighting the need for simplified language, more images, and a dynamic app with less text. In terms of the appearance of the app, most technology professionals found it easy to manage and helpful but only partially agreed on its coverage and attractiveness. These points will be prioritized in future updates.

The Family Health Strategy is an ideal scenario for promoting the use of this tool since the health professionals in these teams aim to expand patient self-care and promote the accountability of care for the user [18]. The app also reinforces the information given by the multidisciplinary team in the hospital, as many women may be tired or focused on their newborn during postpartum visits and did not absorb the orientations well. The positive results related to usability and potential for app use motivate future updates to improve functionality, update content, and add new features.

Limitations
Audio and video content was suggested to be included in the app, but this would cause the app to not be compact and it would move away from the research proposal.

Conclusion
The Towards Motherhood mobile app is deemed to be a valid tool to promote self-care. Through the search in web stores and a literature review, no other app with a similar objective was found. In future updates, additional functions can be integrated into the app, and it can be translated into other languages to cater to a wider range of populations. A summary of this study is presented in Textbox 1.

Textbox 1. Summary table.

What was known about the subject?
• The use of health apps is a growing trend worldwide and is seen as an attractive and facilitating option.

What did this study add to our knowledge?
• This study highlighted the need for postpartum apps as they are currently scarce in web stores, which only offer them for the pregnancy period.
• This study demonstrated the importance of knowledge translation, providing scientific and reliable content in digital environments.
• The development of a multiprofessional technology with a broad vision was also emphasized.

Acknowledgments
This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior–Brasil (CAPES)—Finance Code 001.

The researchers acknowledge and express their gratitude for the participation and support of their colleagues in completing the survey.

All authors declared that they had insufficient or no funding to support open access publication of this manuscript, including from affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided article processing fee (APF) support for the publication of this article.

Conflicts of Interest
None declared.

References
2. Gonçalves BG. Tempo de amor e adaptação: pesquisa participativa para promover a saúde da mulher e do seu filho no pós-parto. Time of love and adaptation: participatory research to promote the health of the woman and her child in the


A Technology-Supported Guidance Model to Increase the Flexibility, Quality, and Efficiency of Nursing Education in Clinical Practice in Norway: Development Study of the TOPP-N Application Prototype

Andréa Aparecida Gonçalves Nes1, BSc, MSc, PhD; Jaroslav Zlamal1, BSc, MSc; Silje Christin Wang Linnerud1, BSc, MSc; Simen A Steindal1,2, BSc, MSc, PhD; Marianne Trygg Solberg1, BSc, MSc, PhD

1Lovisenberg Diaconal University College, Oslo, Norway
2Institute of Nursing, Faculty of Health Studies, VID Specialized University, Oslo, Norway

Corresponding Author:
Andréa Aparecida Gonçalves Nes, BSc, MSc, PhD
Lovisenberg Diaconal University College
Lovisenberggt 15B
Oslo, 0456
Norway
Phone: 47 90703168
Email: andrea.nes@ldh.no

Abstract

Background: The challenges of nursing shortage in the nursing profession and of limited nursing educational capacity in nursing education in clinical practice need to be addressed to ensure supply according to the demand of these professionals. In addition, communication problems among nursing students, nurse educators, and nurse preceptors; variations in the guidance competence of nurse preceptors; and limited overview from nurse educators on nursing students’ clinical practice are common challenges reported in several research studies. These challenges affect the quality of nursing education in clinical practice, and even though these problems have been highlighted for several years, a recent study showed that these problems are increasing. Thus, an approach is required to ensure the quality of nursing education in clinical practice.

Objective: We aimed to develop a guidance and assessment application to meet the challenges reported in clinical practice. The application intended to increase the flexibility, quality, and efficiency of nursing education in clinical practice. Furthermore, it intended to increase interactive communication that supports guidance and ensure structured evaluation of nursing students in clinical practice.

Methods: This study employed a multidisciplinary user-participatory design. Overall, 23 stakeholders from the project team (ie, 5 researchers, 2 software developers, 1 pedagogical advisor, and 15 user representatives [4 educators, 6 preceptors, and 5 students]) participated in a user-centered development process that included workshops, intervention content development, and prototype testing.

Results: This study resulted in the creation of the Technology-Optimized Practice Process in Nursing (TOPP-N) guidance and assessment application for use as a supportive tool for nursing students, nurse preceptors, and nurse educators in clinical practice. The development process included the application’s name and logo, technical architecture, guidance and assessment module, and security and privacy.

Conclusions: This study offers insights into the development of an evidence-based technological tool to support nursing students, nurse preceptors, and nurse educators in clinical practice. Furthermore, the developed application has the potential to meet several challenges reported in nursing education in clinical practice. After a rigorous development process, we believe that the TOPP-N guidance and assessment application prototype is now ready to be tested in further intervention studies.

(JMIR Hum Factors 2023;10:e44101) doi:10.2196/44101
KEYWORDS
clinical practice; guidance application model; nursing students; constructive alignment; metacognition; technological tool; nursing; nursing profession; application; mobile health, eHealth; educator; communication

Introduction

Background
The need for registered nurses (RNs) is increasing worldwide [1,2], and this has been made clear by the advent of the COVID-19 pandemic [3]. In Norway, a shortage of 28,000 nurses is predicted by 2035 [4]. To meet the future demand for RNs, strategies are needed to increase educational capacity and the number of nursing students (NSs) completing the bachelor’s program in regular time with ensured educational quality.

In line with European Union directives [5,6], 50% of the Norwegian nursing education program consists of clinical practice, and due to educational logistics (ie, to avoid displacing educators and students), it is recommended that clinical placements be near educational institutions. This limits educational capacity, because several clinical placements that are far from educational institutions are not being used. Another challenge in nursing education is the current guidance model used in many educational institutions in European countries, including Norway, which cannot ensure that NSs achieve the learning outcomes and the expected quality of clinical practice [7-9]. Therefore, a strategy is urgently needed to increase the number of eligible clinical placements and ensure the learning quality of NSs and their achievement of expected learning outcomes in clinical education.

Challenges Related to the Norwegian Guidance Model

The current Norwegian guidance model for clinical practice involves the following 3 parties: an NS, a nurse preceptor (NP), and a nurse educator (NE). The NP is a RN employed by the health care institution where the NS is present during the clinical practice period (eg, nursing home and hospital), and the NE is employed by an educational institution (ie, the university or university college where the NS receives nursing education). The NP role involves daily face-to-face guidance, follow-up, supervision, and evaluation of the NS. The NE is responsible for ensuring that the clinical practice period provides the NS with optimal learning and a fair assessment of the achieved learning outcomes. This guidance model for clinical practice is described in detail elsewhere [10].

Users experience the current guidance model as fragmented, and it leads to many challenges in clinical practice [7]. A guidance culture and environment are important for NSs to achieve learning outcomes in practical studies [11], but NSs report feelings of isolation and limited cooperation with their NEs and peers [8,12,13]. Furthermore, NSs work with NPs of varying pedagogical competence in clinical guidance, which directly impacts the achievement of their learning outcomes [14].

Both NPs and NSs stress that NPs need to develop their pedagogical competence [7,15]. Thus, a key to overcoming the challenges in clinical practice is to increase NPs’ pedagogical competence and, consequently, guidance skills [16]. Meanwhile, NPs report that insufficient time is the greatest challenge in guiding NSs in clinical practice, as time-consuming preceptorship must be conducted in addition to all the responsibilities and duties of an RN. NEs also find that the current practice model supports only limited contact among NEs, NSs, and NPs, which could result in insufficient oversight of what is happening in clinical practice. Research shows that when a challenge occurs in clinical practice, the NS and NP wait too long to involve the NE, making it harder to solve challenges at an early stage [7]. Although the above-mentioned challenges have been underlined for many years, a recent research study suggested that the challenges related to clinical practice in nursing education have only increased in the past 10 years [17].

Moreover, due to the increasing complexity of modern health care demands [18], critical thinking is a desired outcome in nursing education. Defined as the process of making a reflective judgment [19] about what to believe or do in a given context [20], critical thinking is needed to acquire the expected competence in clinical practice in nursing education.

Communication is the main element of instruction, and digital communication has developed greatly in recent years, facilitating active and remote learning (eg, the use of digital solutions during the COVID-19 pandemic) [21]. Wireless devices, such as smartphones, tablets, and computers, have become an integral part of society and provide access to necessary information and educational tools independent of physical location [22,23]. These devices and the available technology provided the necessary structure for the development of an application aimed at improving communication among NSs, NPs, and NEs in guidance, mitigating the challenges of nursing education in clinical practice.

When developing a technological tool, such as a guidance application, it is recommended that the developers follow an approach that involves all stakeholders (in this case, NSs, NPs, NEs, researchers with expertise in the explored field, and information technology designers and developers) from the early stage (idea generation) to the final stage (evaluation and implementation). However, few studies have followed this recommendation [24]. Furthermore, a recent mixed methods review showed that technological tools tailored to support guidance of NSs in clinical practice that meet the challenges faced by Norwegian educational institutions and stimulate students’ critical thinking are missing [18].

Objective

This study aimed to develop a guidance and assessment application to meet the challenges in the nursing profession and education; improve communication; support guidance; stimulate NSs’ critical thinking; ensure structured evaluation of NSs in clinical practice; and increase the flexibility, quality, and efficiency of nursing education.
Methods

Study Design
This study was performed between September 2019 and March 2020, and employed a multidisciplinary user-participatory design approach [25,26] to ensure that the developed application would be acceptable (ie, well received, suitable, user friendly, and attractive) and designed to fit the needs of users (NSs, NPs, and NEs) and the context of use. The study is part of a larger project to develop, test the effectiveness of, and implement technology-supported guidance to increase the flexibility, quality, and efficiency of nursing education in clinical practice. The study protocol was published prior to the development of the application [10]. The usability, feasibility, and effect of the developed application have been tested, and the results will be presented in future papers.

Recruitment
NSs were recruited through study information published on Canvas (Instructure Inc), a learning management system used by Lovisenberg Diaconal University College (LDUC), while NPs were recruited through an existing cooperation agreement among LDUC, a selected nursing home, and a university hospital in Oslo. NEs were invited to participate in the study by members of the project group. To be eligible to participate in the development of the application, researchers had to be members of the project group, NSs had to have experience of at least one clinical practice, and NPs and NEs had to have experience of guiding NSs in clinical practice. It was important for user representatives to have experience in primary and secondary health care.

Sample
The application development was led by the principal investigator (AAGN), who is an associate professor with long experience as an RN and NE employing active learning methods. The multidisciplinary project group comprised researchers in nursing education with experience of clinical practice mentoring, who also performed the role of NEs (1 doctoral student, 2 associate professors, and 2 professors). Additionally, 1 application developer and 1 designer (from MOSO, a company that cooperated in developing the application), as well as a pedagogical advisor contributed to the application development. The members of the project group included 4 NEs representing all the year units in nursing education, 6 NPs representing clinical practice in primary and secondary health care, and 5 NSs in their second or third year of education (Table 1). The researchers, NEs, NSs, and pedagogical advisor were from LDUC.

Table 1. Members of the project group during the development process.

<table>
<thead>
<tr>
<th>Members and competence/qualification</th>
<th>Value (N=23), n</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Researcher</strong></td>
<td></td>
</tr>
<tr>
<td>Assistant professor</td>
<td>1</td>
</tr>
<tr>
<td>Associate professor</td>
<td>2</td>
</tr>
<tr>
<td>Professor</td>
<td>2</td>
</tr>
<tr>
<td><strong>Nurse educator</strong></td>
<td></td>
</tr>
<tr>
<td>Assistant professor</td>
<td>4</td>
</tr>
<tr>
<td><strong>Pedagogical advisor</strong></td>
<td></td>
</tr>
<tr>
<td>Assistant professor</td>
<td>1</td>
</tr>
<tr>
<td><strong>Nurse preceptor</strong></td>
<td></td>
</tr>
<tr>
<td>Primary health care</td>
<td>6</td>
</tr>
<tr>
<td>Secondary health care</td>
<td>4</td>
</tr>
<tr>
<td><strong>Nursing student</strong></td>
<td></td>
</tr>
<tr>
<td>Second study year</td>
<td>3</td>
</tr>
<tr>
<td>Third study year</td>
<td>2</td>
</tr>
<tr>
<td><strong>Application developer</strong></td>
<td></td>
</tr>
<tr>
<td>Function</td>
<td>2</td>
</tr>
<tr>
<td>Design</td>
<td>1</td>
</tr>
</tbody>
</table>

Ethics Approval
The study was approved by the Institutional Research Board at LDUC and the Norwegian Centre for Research Data (reference number: 338576). Each participant was informed that the meetings would be documented in minutes and the minutes would be used as data to be analyzed in this study. After receiving the information, the participants provided oral informed consent to participate in the development process of the application.

Application Development
The starting point in the development of the application was a concept adopted from a health care intervention created by the
principal investigator (AAGN), in which patients filled out digital electronic reports (e-reports)/diaries and received situational feedback from an RN regarding management of their diabetes [27]. The idea was that the concept of writing e-reports on one’s learning and the concept of communication between users could be transferred to and further developed into a guidance application for use in clinical practice in nursing education.

Before the development started, a contextual inquiry collected information on users’ needs and the requirements for acceptability in nursing homes and hospital units [25]. Target users were involved to elucidate the challenges related to clinical practice from the perspectives of NSs, NPs, and NEs. The development process also used the results of earlier studies on the challenges in guiding NSs in clinical education [12,17,28,29].

Project Management Based on the Spiral Model of Application Development

The spiral model, an approach to software development that uses an adaptive, incremental, and iterative working method in organized multidisciplinary groups, was chosen due to the need for flexibility. This model focuses on addressing risks and involving users in the development process. The project manager is responsible for generating a win-win outcome for users, customers, group members, and concerned stakeholders [30].

The application development process involved the following phases: (1) planning, (2) requirement analysis, (3) design, (4) programming, and (5) testing. In the first 3 phases, weekly workshops were held to discuss the application’s function, content, and process, which were evaluated with the application developers before decisions were made about application functionality (Figure 1). The programming was done by the developers and tested by the user representatives.

Figure 1. The application development process.

Workshops

As part of the contextual inquiry, stakeholders were invited to a daylong workshop to discuss their needs, ideas, requirements for acceptability, and the potential challenges related to the application’s guidance element. Next, we held weekly 2-hour workshops for 3 months. In the first 2 months, the stakeholders were divided into 3 smaller multidisciplinary groups consisting of 1 researcher and at least one NS, one NP, and one NE. Group discussions lasted 1 hour and were led by the researcher. In the next hour, the whole project group met the application developer and application designer to discuss ideas related to content, functions, and design. We pursued an inductive process in which users could suggest ideas on design and content features to further explore users’ requirements for the application.

After the second workshop, the project leader changed the strategy of the work process. Based on input from previous meetings and the mandatory elements of the application’s content and functionality, a draft of the potential design feature, created by AAGN and JZ, was presented to the project group. The mandatory content was based on the NSs’ assessment...
criteria and their expected learning outcomes in clinical practice. The application’s functions aimed to stimulate critical thinking and promote the achievement of the learning outcomes. After the presentation, AAGN and JZ asked the project group members to give their opinions and suggestions for changes and improvements. In each workshop, a topic related to content, functionality, design, or barriers to use was presented and discussed. All the workshops were facilitated by the project leader (first author) in collaboration with other project group members (JZ, MTS, and SAS). Notes were taken on each group discussion.

**Development of the Application’s Name and Promotional Material**

The development of an application is regarded as the development of a new product [31] that requires a brand name. A brand’s primary role is to convey awareness and a favorable impression of the product [32]. An application’s icon should effectively convey meaning, easily communicate the intended function of the application, and enable the user to create associations and meaning [33]. Ideas for the application’s name, logo, icon, and promotional materials were drafted by JZ and AAGN, and then completed by a dedicated graphic designer. The draft was presented to the project group for suggestions, adjustments, and approval.

**Application Functions Based on Constructive Alignment and the Concept of Metacognition**

The constructive alignment principle [34] was adopted in building the application’s learning functions to ensure a connection between learning outcomes, learning activities, and the assessment of clinical practice. The following 3 main phases of the metacognitive cycle were applied: (1) planning and goal setting, (2) applying strategies and monitoring progress, and (3) evaluating and adapting approaches [35]. Metacognition plays a crucial role in the development of critical thinking, so we aimed to support NSs in achieving the clinical practice’s learning outcomes and developing their critical thinking, as RNs who are competent in critical thinking are better prepared to meet the constantly changing and developing challenges of health care [36].

**Application Content Based on Learning Outcomes and Assessment Criteria**

In nursing education, it is commonly challenging to ensure that the learning outcomes for clinical practice have been achieved, and several approaches to assessment have been tried [37-39]. The developed guidance and assessment application employs a validated research-based assessment instrument, the Assessment of Clinical Education (AssCE) [40], whose copyright holders approved its use. The AssCE was developed in Sweden, has been used since 1999, and was originally developed under the general guidelines for Swedish and international requirements [40]. It assesses NSs’ expected learning outcomes during the entire education period.

**Data Collection and Analysis**

The data included recorded notes from the workshops that were initially summarized by the first author, who used rapid analysis [41] to ensure that the material provided essential input for the ongoing development. Rapid approaches to collecting and analyzing data can accelerate an application’s development while maintaining scientific rigor [42]. The data were extracted and compared across the various workshops to identify similarities and differences in the material. Based on the results of constant data analysis, a requirement specification was elaborated for the application, and a prototype was developed.

**Results**

**Overview**

The results presented in this article are limited to the application prototype, including the guidance and assessment module, technical architecture, security, and privacy.

**Application Prototype**

The developed application prototype was intended to make clinical practice studies more flexible through improved communication, a structured evaluation of NSs, and better integration among NPs, NEs, and NSs. Figure 2 shows the workflow processes of the developed application.
Development of the Application’s Name and Promotional Material

The name chosen for the developed application was Technology-Optimized Practice Process in Nursing (TOPP-N). The initial name concept was presented to the project group. The rationale behind the name was that it should highlight the application’s functions and offer a symbol to which a positive meaning could be attributed. In our case, the word topp in Norwegian means “being the best, being on the top” [43]. In English, the equivalent written word is top [44], which phonetically conveys the same meaning as the Norwegian term. The meaning of “being the best, being on the top” also informed the application’s promotional material, including a poster that shows 3 individuals (NS, NE, and NP) holding digital devices while on their way to the “top,” symbolized as a mountain’s summit (Multimedia Appendix 1). The application’s name also denotes that learning in nursing is a process that, in this case, is supported by technology. In developing the visual elements, such as the application’s icon, we focused on reflecting the meaning of top.

The initial concept of the application’s name was accepted by the project group, but several changes were made in the process of developing the icon. When the first idea for the icon was presented to the project group (Multimedia Appendix 2), the group reacted negatively, saying that the icon indicated a “tour planning application” or “hiking application” rather than a guidance application for NSs. The icon was further conceptualized by JZ and AAGN, and a second idea was presented to the group (Multimedia Appendix 3). This time, the initial goal was to symbolize the application’s functions of cooperation and communication, conceptualized as 3 interacting hands. A Facebook poll was created to choose the most liked icon, and the winning draft icon was then submitted to another designer, who created a final version (Multimedia Appendix 4) that shows 3 hands (representing the application’s stakeholders: NSs, NEs, and NPs) surrounding a digital button (representing technology).

Guidance Module

The TOPP-N guidance module’s main functions include NSs’ e-reports and NPs’ feedback to NSs. By completing e-reports,
NSs establish and plan their learning goals, document their experience in clinical practice, and monitor their progress toward their learning outcomes. The e-reports include a planning report (to be completed before the start of the clinical practice day; see Figure 3) and an achievement report (to be completed at the end of the day; see Figure 4). The NSs can evaluate and adapt their approaches on the basis of ongoing experience, learning in clinical practice, and tailored feedback from NPs. The e-reports employ a multiple choice format based on the AssCE [40]. A text field is also available, making it possible to describe and explain a reported situation as necessary (see Textbox 1, point 2.6).

**Figure 3.** Example of a TOPP-N planning report (mobile version screenshot). TOPP-N: Technology-Optimized Practice Process in Nursing.
Textbox 1. Application content in the area of communication competence.

2. TODAY, I HAVE AS A PLAN TO PRACTICE COMMUNICATION

- 2.1 Communication and interaction with patients
  - 2.1.1 Communicates with patients in an engaged manner.
  - 2.1.2 Listens; shows respect and empathy.
  - 2.1.3 Adapts communication to the patient’s needs, for example, in cases of communication difficulties.
  - 2.1.4. Gives the patient adequate room in the dialogue.

- 2.2 Communication with and interaction with next of kin
  - 2.2.1 Communicates with and listens to the viewpoints of family/friends.
  - 2.2.2 Shows respect and empathy.
  - 2.2.3 Creates a dialogue with family/friends and treats their viewpoints with respect.

- 2.3 Cooperation with various authorities within nursing and health care
  - 2.3.1 Communicates, consults, and confers with others.
  - 2.3.2 Ensures continuity in the patient’s chain of care.
  - 2.3.3 Collects, discusses, and critically evaluates relevant information with various authorities and cooperates to ensure appropriate patient care.
  - 2.3.4 Provides correct information to appropriate authorities.

- 2.4 Informs and teaches patients and family/friends
  - 2.4.1 Identifies individual needs.
  - 2.4.2 Organizes and carries out planned instructions.
  - 2.4.3 Adapts information and instructions for self-care.
  - 2.4.4 Provides health-promoting and preventive advice and support.
  - 2.4.5 Follows up on understanding.
  - 2.4.6. Ensures that the patient and family/friends receive coordinated and continuous information and instructions based on their needs and wishes.
  - 2.4.7. Uses various aids and techniques creatively.

- 2.5 Informs and teaches colleagues and students
  - 2.5.1 Demonstrates the ability to seek out and convey information on the patient, situation, and care problems.
  - 2.5.2 Describes his/her own intended educational outcomes.
  - 2.5.3 Teaches and supervises upper secondary students, classmates, or equivalent students.
  - 2.5.4 Critically evaluates information concerning various care issues and conveys it in an interesting manner.
  - 2.5.5 Teaches and supervises with a view to facilitate development and knowledge growth.

- 2.6 Other (Describe which other areas within communication you must work on to achieve the learning outcome related to the current practice period) – Link to the learning outcome for communication for various practice periods

Seek out learning situations in which communication can be challenging, such as patients with impaired language, hearing, another mother tongue, etc.

With regard to the metacognition cycle, completing the planning report increases NSs’ awareness of the learning outcomes and their ability to focus on activities that contribute to achieving them. Completing the achievement report stimulates students to reflect on what they learned that day. The NSs’ e-reports are stored on a secure server administered by MOSO, to which NPs have immediate access. Thus, NPs can enter the metacognitive cycle at any stage through tailored feedback or guidance that contributes to NSs’ development. NPs’ feedback to NSs is informed by the e-reports and the preceptors’ daily experience of NSs and may be delivered either in writing or verbally through an audio file available through the application. NSs and NPs can also communicate directly with each other by text message. This process gives NSs the opportunity to adjust their plans and activities. To ensure that they share a common understanding of guidance needs and expectations, NSs and NPs must evaluate each generated report using a 5-point Likert scale covering the NSs’ need for guidance in the 6 distinct
nursing competence areas shown in the guidance module (Figures 4 and 5).

**Figure 4.** Example of a TOPP-N achievement report (mobile version screenshot). TOPP-N: Technology-Optimized Practice Process in Nursing.

**Figure 5.** Example of nurse preceptor feedback (mobile version screenshot).
NEs can access reports and feedback overviews at any time (Figure 6) by logging into TOPP-N and can support NSs and NPs when necessary. Consequently, the supervision during clinical practice is well documented, providing a better basis for assessing NSs.

**Figure 6.** Two graphs revealing disagreement between a nursing student and a nurse preceptor related to evaluation of the student’s need for supervision (web version screenshot).

Assessment Module

The assessment module (Figure 7) is a digitalization of the AssCE with 21 assessment points under the following 5 main headings: (1) communication and teaching, (2) the nursing process, (3) examinations and treatments, (4) management and cooperation, and (5) professional approach. Each assessment point includes a visual analog scale to identify the target level achieved during clinical practice. The assessment points are also accompanied by explanatory text that corresponds to the 3 levels of goal achievement: “inadequate achievement of goals,” “good achievement of goals,” and “very good achievement of goals.” Detailed information on the AssCE can be found elsewhere [40]. Digitalizing the AssCE enabled NSs, NPs, and NEs to prepare in advance for evaluation meetings and offered flexibility in conducting the meetings (remotely or in person).
The participation of user representatives was essential for achieving the presented results. They stressed that the application had to accommodate NSs’ varied education levels and respective challenges. Moreover, they emphasized that the application should focus not on daily evaluation but on daily guidance and should offer documentation for a fair summative and formative evaluation of clinical practice. NSs sought input on their daily activities through a solution that offered useful, effective, and personalized advice on learning outcomes. NSs also suggested functionalities to record their daily clinical practice activities and to facilitate direct communication among NSs, NPs, and NEs.

**Technical Architecture**

TOPP-N is distributed through official application stores as an application for the iOS and Android systems on digital devices, such as computers, tablets, and smartphones. The application can also be accessed at a dedicated TOPP-N webpage [45]. Technical decisions were executed by the application developers only following discussions with the project leader (the first author) after the leader had conferred with the project group.

**Security, Privacy, and Cost Considerations**

The developed prototype is in accordance with the European General Data Protection Regulations of 2018 [46]. All locally stored information is encrypted. The development of the TOPP-N application prototype costed 600,000 Norwegian crowns (US $60,000), excluding salaries and overhead costs.

**Currently Applied Guidance Model Versus the Guidance Model Supported by TOPP-N**

Figure 8 illustrates how TOPP-N supports the guidance of NSs in clinical practice. The application aims to fill the gaps (address the challenges) that have been identified in clinical practice and to facilitate better integration and interaction among NPs, NEs, and NSs, thus potentially improving cooperation between educational and clinical practice institutions.
Discussion

Principal Findings

This research developed a guidance and assessment application prototype (TOPP-N) to support guidance, improve communication, stimulate NSs’ critical thinking, and ensure their structured evaluation in clinical practice. TOPP-N aims to enhance the flexibility, quality, and efficiency of nursing education. The development process combined evidence-based knowledge and user-centered approaches. To our knowledge, this is the first study combining evidence-based knowledge with user input to inform the development of such an application for clinical practice in nursing education. TOPP-N was developed using iterative and inductive processes through the spiral model of application development [30] and was based on (1) contextual inquiry and co-design processes that gathered input from stakeholders (NSs, NPs, and NEs), (2) application content based on the AssCE [40], and (3) application functionality informed by the metacognition process [35] and constructive alignment principles [34].

In Norway, the current options for a guidance model in clinical practice do not include the use of technology. A recent mixed methods systematic review by our research group found a research gap regarding the use of technology-supported guidance models in nursing education worldwide [18]. TOPP-N contributes to closing this gap and addresses the challenges in clinical practice reported by several research studies [7,17,28].

We highlight the importance of user involvement and evidence-based knowledge in developing such an application and show how TOPP-N can address the challenges associated with clinical practice in nursing education from users’ (NSs, NPs, and NEs) perspectives while also meeting the needs of educational institutions, health institutions, and the society.

Importance of User Involvement

We invited user representatives (NSs, NPs, and NEs) to participate in developing our technology-supported guidance model, an approach that research indicates brings both advantages and challenges [47,48]. One advantage was that our user representatives recognized and confirmed that their challenges in nursing education in clinical practice were identical to those reported in research studies [7,17,28]. Our user representatives also underlined the need for an accessible solution compatible with their existing daily routines that provided positive input and reminders in daily clinical practice.

We experienced that the involvement of user representatives was essential to successfully develop a TOPP-N prototype tailored to the challenges associated with clinical practice in nursing education. By underlining the need for intuitive effective functionalities that would not demand too much time, the user representatives also helped us stay focused on user friendliness. When developing the project description, we were aware that a lack of user involvement in the development process could lead to high attrition rates and low adherence in the use of the application. Research suggests that people stop using technologies that do not meet their needs, requirements, or daily routines [25]. Consequently, it was important to incorporate the needs and requirements of the specific user group as highlighted in the workshops. However, we are aware that user involvement may also bring challenges in application development. van Velsen et al [49] pointed out that user involvement in eHealth design is challenging because the few involved users represent only a fraction of the larger user group, so their input may be biased and limited. Those authors [49] also noted that there is often an overreliance on user input.

In this application development, we experienced a challenge related to the inductive process of development. We observed that the user representatives found the concept of a guidance and assessment application to be somewhat abstract and that they had difficulties in formulating specific detailed suggestions for content and functions. This was the main rationale for...
changing our strategy after the second workshop and presenting a temporary solution with potential design features to the user representatives. This new feedback-based approach stimulated discussion and new suggestions from the stakeholders. Furthermore, after changing the development approach, we found that the stakeholders could now understand the concept of a digital guidance and assessment application and could inductively suggest new functions and improvements. Our experience aligns with the findings of a study that investigated the challenges related to user representatives when taking a co-design approach to developing technological tools (or solutions) [50].

Evidence-Based Knowledge

One of this study’s main goals was to develop a guidance model that continually stimulates NS’s cultivation of critical thinking. To meet this goal, we adopted metacognition as the theoretical approach for developing the application’s functionalities. Critical thinking is stimulated through a continuing reflective process that demands self-monitoring and self-correction, and metacognition has been shown to be effective at stimulating critical thinking in pedagogical interventions [51,52]. The application’s workflow drew upon constructive alignment principles [34] to guarantee that the chosen learning activities helped NSs achieve their learning outcomes and to ensure that the assessment criteria were tailored to the expected learning outcomes.

Quality Assurance in the Learning Process

The TOPP-N application intends to meet the challenges in clinical practice described by several research studies [7,12,17,28] by ensuring good communication between users, structuring the guidance delivered by NPs, and generating an overview of NSs’ clinical practice performance that is available to NSs, NPs, and NEs. In addition, TOPP-N enables NSs to be aware and frequently reminded of the learning outcomes to be achieved in clinical practice by prompting them to plan their day and report their performance daily. Based on the NSs’ reports and the personal daily guidance given through TOPP-N, NPs provide tailored written or verbal feedback that is saved on a server to which NSs and NEs have immediate access. Based on the feedback from NPs, NSs can improve their plan and strategy to achieve the expected learning outcomes for clinical practice, and NEs can support and coach NPs in providing feedback that ensures a pedagogical approach. Another advantage is that, through the application, NPs can always access the expected learning outcomes, enabling them to prioritize the suggestions in the daily guidance to better achieve the goals of clinical practice in nursing education. The described pedagogical process and the application’s documentation of NSs’ clinical practice support quality assurance. To the best of our knowledge, this is the first developed tool that enables documenting the progress of NSs in clinical practice. In 2022, the TOPP-N application received the quality of education award for higher education in Norway awarded by the Ministry of Education and Research (Multimedia Appendix 5).

Students’ Responsibility for Their Own Learning

Because of the theoretical approach taken in developing its functionality, TOPP-N may stimulate an active learning process and enhance NSs’ responsibility for their own learning. The guidance module allows NSs to choose which learning activities to focus on (see Textbox 1) and to gradually work through the expected learning outcomes at their own pace while being guided by NPs. A recently published study underlined NSs’ need for more guidance [8]. In the evaluation module, the digitalization of the AssCE allows all users (NSs, NPs, and NEs) to prepare for evaluation meetings. The digital evaluation form is available to users from their first day in clinical practice, so NSs can progressively record the finished clinical practice activities that indicate their achievement of learning outcomes and can document their need for further guidance. They can also recall what they have planned and done in the guidance module and use the recorded information and feedback in their self-assessments. Progressively documenting their own development empowers NSs to take command of their own learning, provides a valuable opportunity to establish a basis for the direction of the assessment, and greatly influences the results of the clinical practice evaluation.

Flexibility and Transferability to Other Professional Education

As mentioned, the expected global nurse shortage must be addressed [4], but solving this problem is challenging, as European Union regulations stipulate that 50% of nursing education include mandatory clinical practice, yet there are not enough clinical placements near educational institutions [53]. This study aimed to develop a guidance application that could be delivered in a technological format supported by several digital platforms, such as tablets, smartphones, and the web, on both Android and iOS systems. TOPP-N enables follow-up and evaluation with ensured guidance quality, making it possible to use several available clinical placements far from educational institutions, including following up with students in exchange programs. Limited number of clinical placements is the major reason why Norwegian educational institutions are not able to increase their education capacity. Using the TOPP-N application, new clinical placements can be used, which will help solve this problem. In addition, by ensuring educational quality and improving the preceptor’s competence, TOPP-N facilitates mentoring tailored to the NSs’ needs. Through this approach, a higher number of students can finish their education within the expected duration, leading to a higher educational effectivity. TOPP-N has been developed to be easily adaptable to other professional education that includes practice and that may face challenges similar to those in nursing education.

Strengths, Limitations, and Future Directions

This study included a broad range of stakeholders (NSs, NPs, and NEs participating in the project group, as well as researchers and software developers) from the project planning stage through the development process and testing, as recommended by existing research [25]. Mutual learning and shared understanding are core concepts of participatory design, as they ensure mutual respect between stakeholders and enable everyone to take part in the shared decision-making process. Users are
not technology experts and do not necessarily have the language to articulate what they need from an application [50]. Consequently, retaining the same sample of users and giving them adequate knowledge of the development process may have made it easier for the stakeholders to participate actively in application development. It was considered important to the development of TOPP-N to solicit user feedback on functionality, layout, and how the material was presented. However, including new user representatives might have added new perspectives in the development process that we were not able to identify.

A further limitation in the application’s development process is that the relevant stakeholders included NSs and NEs only from LDUC, excluding input from other nursing educational institutions in Norway. Another limitation is the use of the AssCE as the basis of the application’s development, which limits the use of other assessment forms. To use TOPP-N, nursing programs need to adopt the AssCE [49]. Further development of TOPP-N will allow educational institutions to include and use their own evaluation forms in the system.

The development was guided by existing development recommendations, and used a broad range of service design methods and a user-centered design approach to facilitate cocreation, mutual learning, and shared understanding among the stakeholders. Although the process followed a participatory design to increase the likelihood of acceptability, usability, and feasibility, these elements will be tested in future studies. In addition to high user involvement and stakeholder input, the development process was guided by well-established theory and concepts from metacognition [35] and constructive alignment [34]. This enhances the future potential to find positive effects.

Implications for Clinical Practice
TOPP-N is intended to facilitate NSs’ learning processes, ensure quality guidance, and improve communication among NSs, NPs, and NEs. Daily e-reports before and after a shift in clinical practice promote metacognitive strategies that stimulate self-regulated learning and critical thinking. Ongoing feedback is fundamental to NSs’ professional development, as it provides direction and increases their confidence, motivation, and self-esteem. The TOPP-N guidance and assessment application may enhance NPs’ competence in guidance, improve NSs’ learning outcomes in clinical practice, improve the use of resources by enabling remote guidance, and consequently increase nursing education capacity. Better-educated NSs can enhance the quality of care and consequently improve patient safety [54].

Conclusions
The developed guidance and assessment application will enable NSs to complete daily e-reports, receive feedback from NPs and NEs, and evaluate their learning outcomes in clinical practice based on daily mentoring and documentation. It will also enable remote follow-up by NEs, enabling ongoing support of NSs’ learning progress, and prompt involvement when necessary.

This study offers insights into a user-centered approach for the development of an evidence-based guidance and assessment application to ensure the quality of NSs’ clinical practice, providing a practical example of how a technology-supported guidance model can be developed.

It is important to emphasize that developing an application is a constantly evolving dynamic process, so we must continually focus on further development of its content and function. After a rigorous development process, we believe that the TOPP-N guidance and assessment application prototype is ready to be tested in further intervention studies. The project group has performed follow-up intervention studies to test (usability and feasibility) and examine the effects (randomized controlled trial) of the TOPP-N application. The results will be presented in future research articles.

Acknowledgments
This study received financial support from the Agency for International Cooperation and Quality Enhancement in Higher Education and from Lovisenberg Diocesan University College (project number AKTIV-2018/10101). The authors thank the nursing students, nurse preceptors, and nurse educators who participated in this study. We wish to thank Sigrid Elise Naustdal Ille and Edith Roth Gjevjon for their contribution to the application’s development and their input on the manuscript. We also thank Yavuz Oncu for workflow creation and Cristian Adna for the creation of the Technology-Optimized Practice Process in Nursing (TOPP-N) logo and promotional material. Finally, we wish to thank the University of Agder, Oslo University Hospital, and Cathinka Guldberg-senteret Lovisenberg for their cooperation in this study.

Authors’ Contributions
AAGN was responsible for conceptualization, methodology, investigation, formal analysis, writing (original draft, review, and editing), visualization, and project management. JZ contributed to conceptualization, methodology, investigation, visualization, and writing (review and editing). SCWL, SAS, and MTS contributed to conceptualization, investigation, and writing (review and editing).

Conflicts of Interest
None declared.
Multimedia Appendix 1
Final version of the TOPP-N symbol for promotional material. TOPP-N: Technology-Optimized Practice Process in Nursing.

[ PNG File, 832 KB - humanfactors_v10i1e44101_app1.png ]

Multimedia Appendix 2
First version of the TOPP-N symbol. TOPP-N: Technology-Optimized Practice Process in Nursing.

[ PNG File, 293 KB - humanfactors_v10i1e44101_app2.png ]

Multimedia Appendix 3
First version of the TOPP-N icon. TOPP-N: Technology-Optimized Practice Process in Nursing.

[ PNG File, 503 KB - humanfactors_v10i1e44101_app3.png ]

Multimedia Appendix 4

[ PNG File, 189 KB - humanfactors_v10i1e44101_app4.png ]

Multimedia Appendix 5
Quality of education award for higher education in Norway 2022.

[ PNG File, 1206 KB - humanfactors_v10i1e44101_app5.png ]

References


15. STOR VILJE - LITE RESSURSER En kartlegging av rammebetingelser for veiledning av sykepleierstudenter i kommunehelsestjenesten. Norsk Sykepleierforbund. 2018. URL: https://www.nsf.no/Content/3895428/cache=20182205132729/Praksisrapport%20endelig%20mai%202018.pdf [accessed 2023-01-18]


46. General Data Protection Regulation. URL: https://gdpr-info.eu/ [accessed 2023-01-18]


Abbreviations

AssCE: Assessment of Clinical Education
LDUC: Lovisenberg Diaconal University College
NE: nurse educator
NP: nurse preceptor
NS: nursing student
RN: registered nurse
TOPP-N: Technology-Optimized Practice Process in Nursing

©Andréa Aparecida Gonçalves Nes, Jaroslav Zlamal, Silje Christin Wang Linnerud, Simen A Steindal, Marianne Trygg Solberg. Originally published in Jmir Human Factors (https://humanfactors.jmir.org), 03.02.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in Jmir Human Factors, 03.02.2023, is properly cited.
is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Using Clinical Data Visualizations in Electronic Health Record User Interfaces to Enhance Medical Student Diagnostic Reasoning: Randomized Experiment

Lucille Cheng¹, BA; Yalini Senathirajah², PhD

¹School of Medicine, University of Pittsburgh, Pittsburgh, PA, United States
²Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, United States

Abstract

Background: In medicine, the clinical decision-making process can be described using the dual-process theory consisting of the fast, intuitive “System 1,” commonly seen in seasoned physicians, and the slow, deliberative “System 2,” associated with medical students. System-1—type diagnostic reasoning is thought to be less cognitively burdensome, thereby reducing physician error. To date, limited literature exists on inducing System-1—type diagnosis in medical students through cognitive heuristics, particularly while using modern electronic health record (EHR) interfaces.

Objective: In this experimental pilot study, we aimed to (1) attempt to induce System-1—type diagnostic reasoning in inexperienced medical students through the acquisition of cognitive user interface heuristics and (2) understand the impact of clinical patient data visualizations on students’ cognitive load and medical education.

Methods: The participants were third- and fourth-year medical students recruited from the University of Pittsburgh School of Medicine who had completed 1+ clinical rotations. The students were presented 8 patient cases on a novel EHR, featuring a prominent data visualization designed to foster at-a-glance rapid case assessment, and asked to diagnose the patient. Half of the participants were shown 4 of the 8 cases repeatedly, up to 4 times with 30 seconds per case (Group A), and the other half of the participants were shown cases twice with 2 minutes per case (Group B). All participants were then asked to provide full diagnoses of all 8 cases. Finally, the participants were asked to evaluate and elaborate on their experience with the system; content analysis was subsequently performed on these user experience interviews.

Results: A total of 15 students participated. The participants in Group A scored slightly higher on average than those in Group B, with a mean percentage correct of 76% (95% CI 0.68-0.84) versus 69% (95% CI 0.58-0.80), and spent on average 50% less time per question than Group B diagnosing patients (13.98 seconds vs 19.13 seconds, P=.03, respectively). When comparing the novel EHR design to previously used EHRs, 73% (n=11) of participants rated the new version on par or higher (3+/5). Ease of use and intuitiveness of this new system rated similarly high (mean score 3.73/5 and 4.2/5, respectively). In qualitative thematic analysis of poststudy interviews, most participants (n=11, 73%) spoke to “pattern-recognition” cognitive heuristic strategies consistent with System 1 decision-making.

Conclusions: These results support the possibility of inducing type-1 diagnostics in learners and the potential for data visualization and user design heuristics to reduce cognitive burden in clinical settings. Clinical data presentation in the diagnostic reasoning process is ripe for innovation, and further research is needed to explore the benefit of using such visualizations in medical education.

(JMIR Hum Factors 2023;10:e38941) doi:10.2196/38941
Introduction

In medicine, the clinical decision-making process can be described using the dual-process theory, which postulates cognitive processes consist of the fast, intuitive System 1 and the slow, deliberative System 2 [1-5]. Colloquially, System 1 is described as “your gut feeling” and requires minimal cognitive effort due to the use of past experiences and heuristics, whereas System 2 requires significant cognitive effort and can be associated with hypothesis creation and testing [6]. The System-1–type diagnostic reasoning is assumed to take years of experience to develop and is commonly seen in seasoned physicians, whereas the System-2–type diagnostic reasoning is more associated with learners, such as medical students [7].

System-1–type diagnostic reasoning in novices may be possible through the acquisition of cognitive heuristics. One example is Rosby et al [8], who accomplished this by training students to use System-1–type diagnostic reasoning via rapid repeated exposures to training x-rays and showing that this was effective in contrast to longer, fewer exposures. The less cognitively burdensome System-1–type diagnostic reasoning has benefits for patients by allowing physicians to be more present and reducing the potential for mistakes. Human cognitive capacity is limited and prone to error when overtaxed, yet health care systems require physicians to complete efficiently and accurately several, often unrelated, tasks simultaneously [9-11]. Newly developed tools which can provide a “snapshot” of relevant information and live alongside “visualization tools and graphical representations that better synthesize patient information” have been cited as promising approaches moving forward with electronic health records (EHRs) [9,11].

In this experimental study, we aim to further explore the work done by Rosby et al [8] and the potential of data visualizations in EHRs in the two following ways: (1) to induce System-1–type diagnostic reasoning in inexperienced medical students through the acquisition of cognitive user interface heuristics and (2) to better understand the impact of clinical patient data visualizations on students’ cognitive load and medical education.

Methods

Ethical Considerations

This study received approval from the institutional review board of the University of Pittsburgh Human Research Protection Office under STUDY19020169.

Statistical Analysis

Under institutional review board approval from the University of Pittsburgh, the participants recruited were 15 third- and fourth-year medical students who had completed at least one clinical rotation (Table 1). Students were first asked about their experience with existing EHR products, and basic demographic information was collected. They were then randomly assigned to 1 of 2 groups, Group A or Group B. Similarity between the 2 groups was assessed with the Welch 2-tailed $t$ test and ANOVA of the demographic information collected (Table 1). Subsequently, the participants underwent the 3 steps of the study, which were familiarization, training, and testing.

In the familiarization phase, all participants were shown 8 cases based on real patients with clinical information indicative of nonalcoholic fatty liver disease (NAFLD) or metabolic syndrome. The correct diagnosis of each of the patients was one of the following: (1) has metabolic syndrome and NAFLD, (2) has metabolic syndrome and does not have NAFLD, (3) does not have metabolic syndrome and has NAFLD, or (4) does not have metabolic syndrome nor NAFLD. The correct diagnosis for each case was given to the participants, and case information was displayed on a novel EHR user interface featuring a prominent data visualization component (Figure 1).

Next, students in each group were shown 4 cases and asked to provide a correct, full diagnosis (ie, “has metabolic syndrome and NAFLD”) for all 4 cases in a row or 1 trial (Figure 2). Students in Group A were shown each case within a trial up to 30 seconds per case for a total of 4 trials maximum, whereas Group B participants were shown each case within a trial for up to 2 minutes per case for a total of 2 trials maximum. If the students did not correctly diagnose all 4 cases within a trial before maxing out their allotted trial repeats, they would automatically be moved to the testing phase. During the final test phase, all participants were shown all 8 patient cases and asked to provide a full diagnosis of the patient. There was no time limit for either group.

After the study, the participants were asked to evaluate and elaborate on their experience with the novel EHR design. The questions asked included the following: (1) rating ease of system, (2) rating intuitiveness of system, (3) rating usefulness of system, (4) comparing the novel system with past EHRs used based on intuitiveness, (5) strategies used to compete the tasks, and (6) missing features that would have helped the completion of task and areas for improvement. Questions that asked the participants to give a rating or comparison were formatted on a scale of 1-5, where 1 was the low end of the spectrum (ie, very difficult if asking to rate ease of the system) and 5 being the high end of the spectrum (ie, very easy in the aforementioned example). The other questions were open-ended, and the study facilitators encouraged the participants to speak freely in this section. Unbiased follow-up questions eliciting clarification from the participants were occasionally asked. The Welch $t$ test was performed on quantitative values using STATA/SE (StataCorp) and qualitative thematic analysis using the MAXQDA VERBI software was performed on this user interview portion of the study.

KEYWORDS

electronic health record; EHR; System-1–type diagnostic reasoning; type-1 reasoning; diagnostic; diagnosis; user interface; user design; heuristics; medical education; clinical reasoning; reasoning process; data visualization; hGraph; cognitive burden; cognitive load; medical student; medical school
Table 1. Summary table of participant demographics.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Group A (n=8), n (%)</th>
<th>Group B (n=7), n (%)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;25</td>
<td>2 (25)</td>
<td>0 (0)</td>
<td>.08</td>
</tr>
<tr>
<td>&gt;25</td>
<td>6 (75)</td>
<td>7 (100)</td>
<td></td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td>.43</td>
</tr>
<tr>
<td>Male</td>
<td>4 (50)</td>
<td>5 (71)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>4 (50)</td>
<td>2 (29)</td>
<td></td>
</tr>
<tr>
<td><strong>Class year</strong></td>
<td></td>
<td></td>
<td>.74</td>
</tr>
<tr>
<td>MS3&lt;sup&gt;a&lt;/sup&gt;</td>
<td>3 (38)</td>
<td>2 (29)</td>
<td></td>
</tr>
<tr>
<td>MS4</td>
<td>4 (50)</td>
<td>4 (57)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>1 (13)</td>
<td>1 (14)</td>
<td></td>
</tr>
<tr>
<td><strong>Time (hours/week) spent browsing the internet</strong></td>
<td></td>
<td></td>
<td>.98</td>
</tr>
<tr>
<td>&gt;16</td>
<td>3 (38)</td>
<td>3 (43)</td>
<td></td>
</tr>
<tr>
<td>11-15</td>
<td>2 (25)</td>
<td>2 (29)</td>
<td></td>
</tr>
<tr>
<td>&lt;10</td>
<td>3 (38)</td>
<td>2 (29)</td>
<td></td>
</tr>
<tr>
<td><strong>EHR&lt;sup&gt;b&lt;/sup&gt; usage frequency (days/week)</strong></td>
<td></td>
<td></td>
<td>.42</td>
</tr>
<tr>
<td>≥5</td>
<td>5 (63)</td>
<td>6 (86)</td>
<td></td>
</tr>
<tr>
<td>2-4</td>
<td>1 (13)</td>
<td>1 (14)</td>
<td></td>
</tr>
<tr>
<td>&lt;1</td>
<td>2 (25)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td><strong>EHR products used</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cerner</td>
<td>8 (100)</td>
<td>7 (100)</td>
<td>_&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Epic</td>
<td>8 (100)</td>
<td>7 (100)</td>
<td>_</td>
</tr>
<tr>
<td>Other&lt;sup&gt;d&lt;/sup&gt;</td>
<td>3 (38)</td>
<td>4 (57)</td>
<td>.23</td>
</tr>
</tbody>
</table>

<sup>a</sup>MS: medical student.

<sup>b</sup>EHR: electronic health record.

<sup>c</sup>Not applicable.

<sup>d</sup>Centricity, Aria, and Computerized Patient Record System (CPRS).
Figure 1. One of 8 cases participants were familiarized with, then later asked to diagnose (top) compared to the conventional electronic health record (EHR) screen from Epic all students reported previously using (bottom, via emrsystems.net). All cases were displayed on the same user interface which is based on the open-source visualization hGraph, (hgraph.org). The green circle represents the normal range for the parameters shown. The gray "shadow" formed by the linkage of all the values is intended to allow the user to see patterns that may help in future pattern recognition. ALT: alanine transaminase; APRI: AST to platelet ratio index; AST: aspartate aminotransferase; BUN: blood urea nitrogen; Cr: creatinine; DOB: date of birth; DBP: diastolic blood pressure; FIB4: fibrosis-4; HbA1c: hemoglobin A1c; HDL: high-density lipoprotein; K: potassium; LDL: low-density lipoprotein; MELD: model for end-stage liver disease; Na: sodium; NFS: nonalcoholic fatty liver disease fibrosis score; PROMIS29: patient-reported outcomes measurement information system; SBP: systolic blood pressure.
**Results**

A total of 15 medical students participated in the study. Participants in Group A scored slightly higher on average than participants in Group B, with a mean percentage correct of 76% (95% CI 0.68-0.84) versus 69% (95% CI 0.58-0.80) during the final testing portion. While we fail to reject the null hypothesis ($P=.40$), the participants in Group A spent, on average, 50% less time per question than Group B diagnosing patients during the final, time-unlimited testing portion (13.98 seconds vs 19.13 seconds, $P=.03$). A 2-sample equal variance (independent) 2-tailed $t$ test was performed. The difference was found to be significant but inconclusive due to the small sample size (Table 2).

All participants in both groups had previously used Epic and Cerner, and none of the participants in either groups differed significantly in their perceptions of ease of use or usefulness of these EHRs (Table 3). When comparing the study EHR design to previously used EHRs, both groups on average rated the study EHR on par or higher than the existing EHRs (mean score 3.38/5.0 vs 3.71/5.0 for Group A vs B, respectively). Moreover, 73% (n=11) of all participants rated the new version on par or higher than existing EHRs; the ease of use and intuitiveness of this new system rated similarly high.

Qualitative thematic analysis revealed participants across both groups spoke positively about the visual representation data, in particular the ease in quickly assessing a patient’s overview (n=11, 73.3%), the consistency of the user interface layout and reducing number of clicks (n=10, 66.7%), and intuitive color coding (n=8, 53.3%; Table 4). When asked “What strategies did you utilize to help you complete this task?” some participants discussed pattern recognition (n=8, 53.3%) or using the consistency of the user interface (n=10, 66.7%) in combination with their prior clinical training. Areas of improvement for the interface were including numerical values for patient labs to gauge the severity of the condition (n=12, 80.0%) and more clarification around the central “health score” of the patient (n=7, 46.7%).

<table>
<thead>
<tr>
<th>Table 2. Key analysis of the testing portion of the experiment, split by groups. Major results of interest included the accuracy of diagnoses (represented by mean % correct) and the speed of diagnosis (represented by mean time spent per question).</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Testing</strong></td>
</tr>
<tr>
<td>Percentage of correct questions (95% CI)</td>
</tr>
<tr>
<td>Seconds spent per question</td>
</tr>
</tbody>
</table>
Table 3. Interview questions asked and corresponding responses. All questions were “on a scale of 1-5,” where 1 is the low end of the spectrum (i.e., very difficult) and 5 is the high end of the spectrum (i.e., very easy) for questions pertaining to ease of use.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Values, mean (SD)</th>
<th>Group A (n=8)</th>
<th>Group B (n=7)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived ease of use</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Epic</td>
<td>3.38 (0.92)</td>
<td>3.43 (0.53)</td>
<td>.89</td>
<td></td>
</tr>
<tr>
<td>Cerner</td>
<td>2.88 (0.99)</td>
<td>2.43 (0.53)</td>
<td>.29</td>
<td></td>
</tr>
<tr>
<td>Study EHR</td>
<td>3.63 (0.92)</td>
<td>3.86 (1.07)</td>
<td>.66</td>
<td></td>
</tr>
<tr>
<td>Perceived usefulness</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Epic</td>
<td>4.38 (0.74)</td>
<td>4.29 (0.76)</td>
<td>.82</td>
<td></td>
</tr>
<tr>
<td>Cerner</td>
<td>4.25 (0.89)</td>
<td>3.86 (0.90)</td>
<td>.41</td>
<td></td>
</tr>
<tr>
<td>Study EHR</td>
<td>3.75 (0.71)</td>
<td>3.57 (0.79)</td>
<td>.65</td>
<td></td>
</tr>
<tr>
<td>Perceived intuitiveness</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study EHR</td>
<td>4.13 (0.83)</td>
<td>4.29 (1.11)</td>
<td>.76</td>
<td></td>
</tr>
<tr>
<td>Ease of use comparison of study EHR to familiar EHR</td>
<td>3.38 (1.30)</td>
<td>3.71 (1.11)</td>
<td>.59</td>
<td></td>
</tr>
</tbody>
</table>

*aEHR: electronic health record.

Table 4. Key themes and choice quotes from the participants based on MAXQDA analysis.

<table>
<thead>
<tr>
<th>Theme</th>
<th>Frequency</th>
<th>Quote</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ability to quickly assess a patient’s overall health</td>
<td>11</td>
<td>“I really like the wheel concept because you’re getting a picture of every component of the patient’s health” (Group A participant)</td>
</tr>
<tr>
<td>Consistency of interface layout and reduced number of clicks aiding ease of use</td>
<td>10</td>
<td>“This was a lot more intuitive than [other EHR] where it’s just a bunch of abnormal labs you have to double click to see if it’s high or low” (Group B participant)</td>
</tr>
<tr>
<td>Intuitive color-coding aiding ease of use</td>
<td>8</td>
<td>“Once I got used to it…visually, it was very easy to see that a bright orange cluster was a [metabolic syndrome] cluster” (Group B participant)</td>
</tr>
<tr>
<td>Pattern recognition as a strategy used to accomplish task</td>
<td>8</td>
<td>“It felt very natural to look at the right areas…after a few patients, my eyes were moving where they needed to go” (Group A participant)</td>
</tr>
<tr>
<td>Desire to have numerical lab values included to gauge the severity of the patient’s condition</td>
<td>12</td>
<td>“It was nice to see trends over times, but without a number I don’t know what the patient’s baseline is…” (Group A participant)</td>
</tr>
<tr>
<td>Confusion around central “health score” (i.e., large number in the middle of the data visualization)</td>
<td>7</td>
<td>“I knew [the number] was important, but I didn’t know what information it was conveying” (Group B participant)</td>
</tr>
</tbody>
</table>

*aEHR: electronic health record.

Discussion

Principal Results

In this study, we attempted to induce System-1 diagnostic reasoning in medical students by using a novel EHR data visualization design. Despite the failure to reject the null hypothesis, we observed a statistically significant difference in the amount of time Group A participants took to fully diagnose patient cases compared with Group B. The increased speed of diagnosis is a key component in System-1-type diagnostic reasoning, as physicians are presumed to rely on pattern recognition based on their past experiences and heuristics as opposed to exerting cognitive effort on the spot. This finding, coupled with the trending results of more accurate diagnoses by Group A than Group B, is suggestive of the ability to induce an accurate System-1-type clinical diagnostic reasoning ability in medical students using frequent repeat exposures. This is akin to the findings by Rosby et al [8].

For students to successfully accomplish the given task of diagnosing whether a patient had NFLD or metabolic syndrome all in a few minutes, several spoke about using “pattern-recognition” cognitive heuristic strategies consistent with System 1 decision-making. These patterns generally fell into one of the following three categories: (1) consistency of layout aiding in finding specific lab values, (2) trends between different lab values and subsequent diagnosis, and (3) visualization-specific features such as color coordination. One participant spoke of the “search pattern” they had developed through medical school and believed was represented in the layout of the user interface, stating the following:

the way the page was set up, it felt very natural to look at the right areas. I would look at BMI first, then down at ALT and AST [common lab values for
diagnosing metabolic syndrome and non-alcoholic fatty liver disease]. …after a few patients, my eyes were moving where they needed to go.

Limitations
Our study had several limitations. First, it was only performed with 15 students at 1 academic institution, thereby making generalizability unlikely. Additionally, only 1 data visualization interface was shown to all participants. The specific design that was used may not adequately represent other potential iterations of clinical data visualizations on EHRs and again makes generalizability unlikely.

We also did not compare the efficacy of the novel EHR with an existing EHR interface such as Epic or Cerner, as our participant criteria included previous EHR experience, and we were interested in the ability to induce System-1 thinking with a completely novel system. We chose to limit our study to participants who had previously used some sort of EHR as these participants were able to provide us design feedback informed by their past clinical experience, as opposed to purely aesthetic feedback on the novel EHR design.

Finally, the incorporation of data visualization into EHRs is limited to the decision of the EHR companies; while there may be some benefit to teaching students clinical data through more illustrative methods, this benefit may be moot if visualizations are not adopted on the primary platform where students perform their clinical duties.

Further Considerations
Several of the qualitative themes hold promise in further investigation of amalgamating the current offerings of how medical education is delivered. Many of the issues students mentioned with current EHRs are solved usability problems in the consumer technology industry by companies such as Google and Apple, but the solutions are not widely adopted in health care today [12,13]. Similarity, the value of data visualizations is not new [14-16], but to our knowledge, this type of clinician-side data visualization is not widely used in medical education.

The heuristics participants alluded to mirror the widely accepted “10 Usability Heuristics” in consumer user experience web design by Nielsen [17], or foundational principles established by Nielsen in 1994 for evaluating the usability of website interfaces [18]. We will focus on the following 2 in particular: Heuristic #4, or “Consistency and Standards,” as well as Heuristic #8, “Aesthetic and Minimalist Design.”

We begin first by talking about Heuristic #4, which states that “users should not have to wonder whether different words, situations or actions mean the same thing. Follow platform and industry conventions” [19]. When taken into consideration with the Jakob Law, or the fact that most users are spending their time on products other than EHRs; introducing a new interface through an EHR that works differently from the consumer products users are accustomed to increases cognitive load by forcing them to learn something new [20]. Data visualizations and illustrative representations of data have become increasingly common user interfaces in consumer technology products such as Jawbone UP and Fitbit [21]. The efficacy of modeling the novel EHR interface after these known patterns was reflected in the higher-than-average intuitiveness scores given by most participants. One noted that the interface “looks like something you’d show a patient…like it would be on the front page of [the patient facing hospital account portal].” This serves as a good reminder that medical students, in addition to becoming physicians, are patients and technology consumers who have to context switch every time they use present-day EHRs.

Heuristic #8, Aesthetic and Minimalist Design, builds upon the basis set by Heuristic #4 and states, “interfaces should not contain information which is irrelevant or rarely needed. Every extra unit of information in an interface competes with the relevant units of information and diminishes their relative visibility” [19]. The novel EHR interface shown to participants was built on hGraph, whose creators were inspired to reduce problems in the health care experience resulting from an excess of data. They did so by using the “single picture method,” which compiled multiple metrics into a unified graph with the belief that “healthcare information visualizations should enable pattern recognition” [21]. In our study, participants spoke to this inadvertently through their comments about the ability to quickly assess a patient’s overall health. In total, 73% of the participants appreciated the ability to easily see an overview of their patients’ health and intuited that they would be able to get more details in a more interactive version of the interface. These heuristics are especially important considering the influence technology has had on our participants’, and broadly, current millennial medical students’ visual consumption of content [22].

Conclusions
How clinical data are presented in the diagnostic reasoning process and medical education is ripe for innovation. In this study, students were able to diagnose patients more accurately after short, repeated exposure to the data visualization interface, implying the possibility of inducing type-1 diagnostics. Additionally, this study demonstrates how incorporating data visualizations and user design heuristics during care delivery can potentially reduce cognitive burden and allow even novices to diagnose quickly and correctly. Further experiments on different, visual displays of data and the benefits it may have on medical education should be conducted, especially in comparison to the existing commonly used EHRs. Studies using eye tracking to better understand what patterns students used, as well as which features were most or least used, should also be run to more precisely understand the search patterns mentioned by the students.

Acknowledgments
The authors wish to thank GoInvo, the team behind the original hGraph product, as well as the University of Pittsburgh Department of Biomedical Informatics and School of Medicine for funding research efforts.
Conflicts of Interest
None declared.

References
16. Cawthon N, Moere AV. The Effect of Aesthetic on the Usability of Data Visualization. 2007 Presented at: 11th International Conference Information Visualization (IV '07); July 04-06, 2007; Zurich, Switzerland.
19. Follert J, Sonin J. hGraph: An open system for visualizing personal health metrics. hGraph. URL: https://www.hgraph.org/docs/hGraph_Whitepaper.pdf [accessed 2023-03-10]
22. Which language do the Millennials understand? The one that starts with V. Infographic Design Team. 2020 Apr. URL: https://www.infographicdesignsteam.com/blog/which-language-do-the-millennials-understand-the-one-that-starts-with-v/ [accessed 2023-03-10]

Abbreviations
- EHR: electronic health record
- NAFLD: nonalcoholic fatty liver disease
User Intentions to Use ChatGPT for Self-Diagnosis and Health-Related Purposes: Cross-sectional Survey Study

Yeganeh Shahsavar1*, MSc; Avishek Choudhury1*, PhD

Industrial and Management Systems Engineering, Benjamin M Statler College of Engineering and Mineral Resources, West Virginia University, Morgantown, WV, United States

*all authors contributed equally

Corresponding Author:
Avishek Choudhury, PhD
Industrial and Management Systems Engineering
Benjamin M Statler College of Engineering and Mineral Resources
West Virginia University
1306 Evansdale Drive
321 Engineering Sciences Building
Morgantown, WV, 26506
United States
Phone: 1 3042934970
Email: avishek.choudhury@mail.wvu.edu

Abstract

Background: With the rapid advancement of artificial intelligence (AI) technologies, AI-powered chatbots, such as Chat Generative Pretrained Transformer (ChatGPT), have emerged as potential tools for various applications, including health care. However, ChatGPT is not specifically designed for health care purposes, and its use for self-diagnosis raises concerns regarding its adoption’s potential risks and benefits. Users are increasingly inclined to use ChatGPT for self-diagnosis, necessitating a deeper understanding of the factors driving this trend.

Objective: This study aims to investigate the factors influencing users’ perception of decision-making processes and intentions to use ChatGPT for self-diagnosis and to explore the implications of these findings for the safe and effective integration of AI chatbots in health care.

Methods: A cross-sectional survey design was used, and data were collected from 607 participants. The relationships between performance expectancy, risk-reward appraisal, decision-making, and intention to use ChatGPT for self-diagnosis were analyzed using partial least squares structural equation modeling (PLS-SEM).

Results: Most respondents were willing to use ChatGPT for self-diagnosis (n=476, 78.4%). The model demonstrated satisfactory explanatory power, accounting for 52.4% of the variance in decision-making and 38.1% in the intent to use ChatGPT for self-diagnosis. The results supported all 3 hypotheses: The higher performance expectancy of ChatGPT (β=.547, 95% CI 0.474-0.620) and positive risk-reward appraisals (β=.245, 95% CI 0.161-0.325) were positively associated with the improved perception of decision-making outcomes among users, and enhanced perception of decision-making processes involving ChatGPT positively impacted users’ intentions to use the technology for self-diagnosis (β=.565, 95% CI 0.498-0.628).

Conclusions: Our research investigated factors influencing users’ intentions to use ChatGPT for self-diagnosis and health-related purposes. Even though the technology is not specifically designed for health care, people are inclined to use ChatGPT in health care contexts. Instead of solely focusing on discouraging its use for health care purposes, we advocate for improving the technology and adapting it for suitable health care applications. Our study highlights the importance of collaboration among AI developers, health care providers, and policy makers in ensuring AI chatbots’ safe and responsible use in health care. By understanding users’ expectations and decision-making processes, we can develop AI chatbots, such as ChatGPT, that are tailored to human needs, providing reliable and verified health information sources. This approach not only enhances health care accessibility but also improves health literacy and awareness. As the field of AI chatbots in health care continues to evolve, future research should explore the long-term effects of using AI chatbots for self-diagnosis and investigate their potential integration with other digital health interventions to optimize patient care and outcomes. In doing so, we can ensure that AI chatbots, including ChatGPT, are designed and implemented to safeguard users’ well-being and support positive health outcomes in health care settings.
KEYWORDS
human factors; behavioral intention; chatbots; health care; integrated diagnostics; use; ChatGPT; artificial intelligence; users; self-diagnosis; decision-making; integration; willingness; policy

Introduction

Background

The digital age has witnessed an unprecedented surge in technological innovation, shaping the essence of human-computer interaction. As the world progresses toward a future encompassing artificial intelligence (AI), advanced conversational AI models, such as Chat Generative Pretrained Transformer (ChatGPT), a cutting-edge conversational AI model by OpenAI, have come to the forefront of academic discussion. This paradigm-shifting technology has revolutionized our interactions with machines and introduced profound implications across multiple disciplines. By harnessing the power of machine learning, ChatGPT transcends the limitations of traditional chatbots, yielding increasingly humanlike conversational capabilities. The technology has demonstrated remarkable capabilities, such as understanding context, generating coherent text, and adapting to various natural language processing (NLP) tasks, including but not limited to language translation, answering of questions, and text generation [1]. The success of these models can be attributed to their scale, as they have been trained on vast amounts of data from diverse sources, such as books, papers, and websites [2]. By leveraging these extensive training data, ChatGPT has learned patterns, syntax, and semantics, enabling it to generate humanlike responses, making it a valuable tool in many applications and industries [3].

The literature has demonstrated the potential of AI-based chatbots, such as ChatGPT, to revolutionize patient care and service delivery [4-6]. Numerous recent studies have underscored the potential of ChatGPT in the health care sector [7]. For instance, a study examined the capabilities of ChatGPT across a range of clinical situations, discovering its potential to enhance patient communication and engagement within health care contexts. The study found that ChatGPT effectively delivers information and support to patients in various scenarios, such as mental health assessments, counseling, medication management, and patient education [8]. A recent review examined the advantages of ChatGPT and other large language models in augmenting medical education, streamlining clinical decision-making, and promoting better patient outcomes [9].

However, ChatGPT is not specifically trained in medical literature. It is crucial to understand the intended purpose of ChatGPT and acknowledge its negative consequences if used otherwise. The use of ChatGPT in health care has raised concerns about the accuracy and reliability of the information provided, patient privacy and data security, prejudice, responsibility, and the ethical ramifications of using such potent language models. A study also emphasized the importance of using strong cybersecurity measures to protect patient data and privacy when using ChatGPT in health care settings [10].

Although ChatGPT has proven to be a remarkable technological achievement, its application in self-diagnosis poses significant risks that must be noted. We all have used the internet for self-diagnosis. Depending on the user’s health literacy, the source’s validity, and the accuracy of information interpretation, web-based self-diagnosis has resulted in positive and negative consequences. Just like most consumer-facing screen-based technologies, ChatGPT has the potential for misinterpretation and misuse, necessitating a careful approach to implementation. This is important because misuse (using it for tasks it is not designed for) can affect user trust, resulting in the underuse of the technology [11]. The convenience and accessibility of ChatGPT have made it appealing for self-diagnosis purposes, much like the broader internet. With an internet connection, ChatGPT can be easily accessed anytime and anywhere, allowing individuals to seek diagnostic information without needing a physical appointment or incurring medical costs. This ease of access can be incredibly enticing for those with limited access to health care services or who face financial constraints. Another factor contributing to the appeal of ChatGPT for self-diagnosis is the sense of anonymity and privacy it provides. Discussing sensitive health issues can be uncomfortable or embarrassing, leading individuals to prefer the discretion offered by an AI-based chatbot over face-to-face consultations with health care professionals. ChatGPT delivers prompt responses, providing instant feedback to users’ inquiries. This immediacy can attract those seeking quick answers or reassurance about their health concerns. Additionally, as ChatGPT is an AI-driven language model built on vast amounts of data and because of its promising performance in several fields, users may perceive it as a knowledgeable and reliable source of information. This perceived expertise can create a false sense of confidence in the diagnostic suggestions provided by ChatGPT, despite its inherent limitations.

We must understand that anyone with a computer and an internet connection can use ChatGPT. Individuals with minimal to no health and technology literacy may not realize the limitations of ChatGPT and its intended use. User character, the intricacy of medical information, and the unique nature of individual patient cases underscore the potential for misinterpretation. Inaccurate or incomplete information provided by ChatGPT may result in misguided self-diagnosis or exacerbation of existing conditions. From a cognitive human factor standpoint, the misalignment between AI-generated information and users’ mental models can lead to erroneous decision-making and unfavorable health outcomes. The possibility of ChatGPT being misused for self-diagnosis is a significant concern. To counteract this, accessing the potential misuse of ChatGPT from a human factor standpoint is essential.

Addressing the concerns associated with AI chatbots in health care, this study aims to (1) explore users’ intentions to use ChatGPT for self-diagnosis and (2) gain a deeper understanding...
of the factors influencing their decision-making processes. By providing novel insights into the implications of AI chatbot adoption in health care settings, we intend to inform the development of guidelines, policies, and interventions that promote the responsible and effective use of AI technologies in health care. The originality of this study stems from its focus on users’ decision-making processes and intentions when using AI chatbots for self-diagnosis, an area of research that remains relatively unexplored in the context of health care applications.

As illustrated in Figure 1, our investigation examines the effects of the perceived effectiveness and risk-benefit appraisal of ChatGPT on decision-making and the subsequent impact on users’ intent to use ChatGPT for self-diagnosis. This examination is crucial due to AI technologies’ rapid growth and adoption across various aspects of daily life, including health care and self-diagnosis [12]. Gaining a comprehensive understanding of the factors driving user acceptance, trust, and adoption of AI technologies is essential to ensure their responsible and efficient use. Additionally, scrutinizing the potential implications of these effects is critical for informing guidelines and policies surrounding AI technologies like ChatGPT for self-diagnosis [13].

By pinpointing the factors contributing to users’ decision-making processes and intentions to use ChatGPT, regulators and health care professionals can develop informed policies and recommendations to ensure AI’s safe and ethical deployment in health care [14]. Adopting this approach will help mitigate potential misuse or overreliance on such technologies for self-diagnosis, which could result in misdiagnosis or delayed treatment.

**Figure 1.** Conceptual framework illustrating the effect of performance expectancy and risk-reward appraisal on the perception of decision-making (directly) and intent to use ChatGPT for self-diagnosis (indirectly). ChatGPT: Chat Generative Pretrained Transformer; H: hypothesis.

**Theory and Hypotheses Development**

The conceptual framework (see Figure 1) explored in this study was inspired by the unified theory of acceptance and use of technology (UTAUT). UTAUT is an established theoretical framework extensively used for comprehending and predicting individuals’ technology adoption and usage [15]. The UTAUT framework posits 4 factors influencing an individual’s behavioral intention to use a given technology: performance expectancy, effort expectancy, social influence, and facilitating conditions. In this research, we only retrieved performance expectancy from UTUAT and added risk-benefit considerations and decision-making as factors affecting the intent to use ChatGPT.

**Operational Definitions**

In this study, performance expectancy is operationally defined as the extent to which an individual anticipates that using ChatGPT will augment their capacity to accomplish tasks, attain objectives, and alleviate workload proficiently and efficiently. This latent construct encapsulates the user’s perceptions regarding the advantages, effectiveness, and overall satisfaction derived from their interaction with ChatGPT.

The construct of decision-making was operationally defined as the extent to which an individual perceives ChatGPT as a valuable tool for assisting them in making informed, timely, and effective choices by providing relevant recommendations and insights. This latent construct encompasses the user’s belief in ChatGPT’s ability to contribute positively to their decision-making process and their willingness to act on the recommendations generated by the technology.

Similarly, the risk-reward-appraisal construct can be operationally defined as the extent to which an individual perceives the advantages of using ChatGPT as surpassing any potential adverse consequences or risks associated with its use. This latent construct captures the user’s evaluation of the trade-offs between the positive outcomes derived from ChatGPT and the potential hazards or drawbacks that may arise from its implementation.

**Hypotheses**

The following are the hypotheses tested in this study.

**Hypothesis 1**

The higher performance expectancy of ChatGPT is positively associated with improved user decision-making outcomes. Hypothesis 1 (H1) is grounded in established theories, such as the technology acceptance model (TAM) and UTAUT [15,16]. These theories posit that performance expectancy is critical to technology acceptance and usage intentions. Trust in a technology, which is positively associated with performance expectancy [17], further supports the notion that when users have higher trust in ChatGPT’s ability to perform effectively, they are more likely to rely on its recommendations, thus positively influencing their decision-making processes [11]. Cognitive fit theory complements this relationship by suggesting that the alignment between an individual’s cognitive processes and the representation of information by technology influences the effectiveness of problem solving and decision-making [18].
As users perceive ChatGPT as an effective tool that aligns with their cognitive processes and expectations, they are more inclined to incorporate its recommendations into their decision-making, leading to improved outcomes.

Although this is the first study to explore the impact of the perceived effectiveness of ChatGPT on decision-making, H1 can be justified by drawing on several studies that have explored the relationship between performance expectancy and technology acceptance, usage, or decision-making in other domains. For instance, Al-Emran et al [19] conducted a systematic review investigating the impact of performance expectancy on mobile learning adoption, revealing its positive effect on learners’ intentions to use mobile technologies for educational purposes. This study further emphasizes the importance of performance expectancy in shaping users’ engagement with technology and their inclination to rely on it for decision-making. Lee and Kozar [20] explored the relationship between website quality, user satisfaction, and decision-making in e-business contexts. Their findings demonstrated that when users perceive a website as effective and efficient, they are more likely to trust its recommendations and make decisions based on the provided information. This study underscores the significance of performance expectancy in users’ trust and decision-making behaviors. Another study proposed that managers’ perceived usefulness and ease of use of AI are significant predictors of their intention to use AI for decision-making in organizations [21]. In a study by Aliaid and Zhou [22], the determinants of health care professionals’ intention to adopt AI-based clinical decision support systems were examined, focusing on factors such as performance expectancy, effort expectancy, social influence, and facilitating conditions. The researchers proposed an extended TAM tailored for health care, addressing the distinct requirements and challenges of the health care domain. The study offered insights into the factors that affect health care professionals’ decision-making processes and their intent to use AI technologies in their practice [22].

Hypothesis 2
A positive risk-reward appraisal of ChatGPT is associated with enhanced user decision-making outcomes. Hypothesis 2 (H2) is based on established psychological and decision-making theories, such as prospect theory and protection motivation theory (PMT), as well as the concept of trust in technology [17,23,24].

Prospect theory posits that individuals evaluate potential gains and losses during decision-making processes and that their choices are influenced by the perceived risks and rewards associated with each alternative [23]. In the context of ChatGPT, users who perceive the benefits of using AI technology to outweigh the potential risks are more inclined to rely on it for decision-making purposes. PMT suggests that individuals’ intentions to engage in protective behaviors are influenced by their perceived severity of a threat, perceived vulnerability, response efficacy, and self-efficacy [24]. Applying PMT to ChatGPT implies that if users believe the benefits of using the technology (response efficacy) surpass the potential risks (perceived severity and vulnerability), they are more likely to integrate ChatGPT’s recommendations into their decision-making processes. Moreover, trust in technology has been identified as a crucial factor influencing technology adoption and usage [17]. When users perceive a favorable risk-reward balance, they are more likely to trust ChatGPT and subsequently rely on its recommendations for decision-making.

In technology adoption, risk perception can significantly affect decision-making. Although there may not be studies directly examining the relationship between risk-reward appraisal and decision-making in the context of ChatGPT, several studies have explored the impact of risk perception and trust in technology on decision-making and technology adoption in other domains. For instance, a study examined the interplay between trust, perceived risk, and TAM in the context of consumer acceptance of electronic commerce (e-commerce) [25]. The findings revealed that trust and perceived risk significantly influence users’ behavioral intentions. Users who perceived a favorable risk-reward balance were more inclined to engage with e-commerce platforms [25]. This suggests that a positive risk-reward appraisal could also influence decision-making processes involving ChatGPT. Another study developed a trust-based consumer decision-making model in e-commerce, emphasizing the importance of perceived risk and trust in users’ decision-making processes [26]. The study demonstrated that users who perceive a positive risk-reward balance when using e-commerce platforms are more likely to base their decisions on the information provided, further supporting the notion that risk-reward appraisal plays a crucial role in decision-making outcomes [26]. Lastly, a study explored the role of trust and risk perception in mobile commerce adoption. Their findings indicated that users who perceive a favorable risk-reward balance are likelier to adopt mobile commerce technologies and rely on them for decision-making [27]. This study highlights the significance of risk-reward appraisal in technology adoption and decision-making.

Hypothesis 3
A positive perception of ChatGPT’s role in enhancing decision-making processes is associated with an increased intention among users to use the technology for self-diagnosis. Hypothesis 3 (H3) can be substantiated by drawing upon well-established theories, such as TAM, UTAUT, and research on trust in technology [15-17]. TAM posits that users’ intention to adopt a technology is influenced by their perceptions of its usefulness and ease of use [16]. Consequently, if users view ChatGPT as a valuable decision-making aid that is user friendly, they are more likely to intend to use it for self-diagnosis purposes. TAM also asserts that users’ actual system usage is impacted by their behavioral intention, suggesting that positive decision-making experiences with ChatGPT could lead to increased use for self-diagnosis. According to UTAUT, when users experience effective decision-making processes that involve ChatGPT (technology), their performance expectancy (the extent to which they believe the technology will assist them in achieving their goals) may rise, thereby fostering a greater intent to use ChatGPT for self-diagnosis.
Methods

Ethical Considerations
The study, classified as a flex protocol type, received approval from the Institutional Review Board (IRB) of West Virginia University (IRB protocol number 2302725983). Informed consent was obtained from participants. The data gathered through the online survey were securely stored on Centiment’s platform and remained accessible exclusively to the research team.

Survey Instruments
Table 1 lists the survey questions used in the study. We adapted questions from UTAUT to form the latent construct performance expectancy. The construct was established by aggregating questions related to 4 statements:

- Statement 1: “ChatGPT can help me achieve my goals.” This item gauges the user’s conviction regarding ChatGPT’s capability to facilitate the attainment of their desired objectives within the context of their tasks.
- Statement 2: “ChatGPT can reduce my workload.” This item appraises the user’s perception of ChatGPT’s potential to mitigate the burden of task completion by streamlining processes and increasing efficiency.
- Statement 3: “I was successful in achieving what I wanted to accomplish with ChatGPT.” This item measures the user’s perception of the degree to which their interaction with ChatGPT has led to the realization of intended outcomes, reflecting the efficacy of the technology in practical applications.
- Statement 4: “I am satisfied with ChatGPT.” This item examines the user’s overall contentment with the performance of ChatGPT, capturing their appraisal of its utility and effectiveness in addressing their needs and expectations.

In addition, 2 statements were developed to form the latent construct decision-making:

- Statement 1: “ChatGPT helps me make informed and timely decisions.” This item evaluates the user’s perception of ChatGPT’s capacity to provide pertinent information, insights, and guidance, which in turn facilitates well-informed and timely decision-making processes.
- Statement 2: “I am willing to make decisions based on the recommendations provided by ChatGPT.” This item measures the user’s trust in the recommendations offered by ChatGPT and their readiness to incorporate those suggestions into their decision-making processes.

Table 1. Statements used in the survey.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance expectancy (PE)</td>
<td>To what extent do you agree with the following: ChatGPT(^a) can help me achieve my goals (PE1).</td>
</tr>
<tr>
<td></td>
<td>To what extent do you agree with the following: ChatGPT can reduce my workload (PE2).</td>
</tr>
<tr>
<td></td>
<td>To what extent do you agree with the following: I was successful in achieving what I wanted to accomplish with ChatGPT (PE3).</td>
</tr>
<tr>
<td></td>
<td>To what extent do you agree with the following: I am satisfied with ChatGPT (PE4).</td>
</tr>
<tr>
<td>Perception of decision-making (DM)</td>
<td>To what extent do you agree with the following: ChatGPT helps me make informed and timely decisions (DM1).</td>
</tr>
<tr>
<td></td>
<td>To what extent do you agree with the following: I am willing to make decisions based on the recommendations provided by ChatGPT (DM2).</td>
</tr>
<tr>
<td>Risk-reward appraisal (RRA)</td>
<td>To what extent do you agree with the following: The benefits of using ChatGPT outweigh any potential risks (RRA).</td>
</tr>
<tr>
<td>Intent to use (IU)</td>
<td>To what extent do you agree with the following: I am willing to use ChatGPT for self-diagnosis purposes (IU).</td>
</tr>
</tbody>
</table>

\(^a\)ChatGPT: Chat Generative Pretrained Transformer.

Furthermore, 1 statement measured the extent to which an individual perceives the advantages of using ChatGPT as surpassing any potential adverse consequences or risks associated with its use: “The benefits of using ChatGPT outweigh any potential risks.”

Lastly, users’ willingness to use ChatGPT for self-diagnosis was captured using 1 statement: “I am willing to use ChatGPT for self-diagnosis purposes.”

All the items were assessed using a 4-point Likert scale, allowing participants to indicate their level of agreement with each statement, ranging from “strongly disagree” to “strongly agree.”

Note that we used a forced Likert scale in this study. By precluding the inclusion of a neutral or midpoint option, a forced Likert scale necessitates respondents to articulate a definitive opinion or predilection, thereby generating data that are more incisive and unequivocal [28]. This approach proves particularly advantageous in scenarios where the research is intended to ascertain well-defined attitudes or perceptions from participants. Forced scales have been demonstrated to mitigate the acquiescence bias, a phenomenon wherein respondents are predisposed to concur with statements regardless of their content [29]. Eliminating a neutral option encourages participants to critically contemplate their responses, yielding more accurate data [28]. Furthermore, forced scales engender more reliable...
outcomes when assessing relatively polarized or fervently held attitudes [30]. By obliging participants to select between affirmative and negative response options, a forced scale can offer more lucid insights into the direction and intensity of their attitudes.

Data Collection
The data collection for this study took place in February 2023, using an online survey administered through Centiment, a reputable service provider for survey deployment and data gathering [31]. By leveraging Centiment’s capabilities, the research team efficiently disseminated the survey. It ensured the participation of a diverse sample of respondents, specifically recruiting individuals who used ChatGPT at least once per month. Centiment’s robust platform facilitated the research team in designing and distributing the survey, while implementing various quality control measures and preventing duplicate responses. This approach bolstered the data’s reliability and validity. Furthermore, the online survey format allowed participants to complete it at their discretion, contributing to an increased response rate and enhanced sample diversity.

Upon obtaining informed consent from participants, they were directed to the survey, which contained questions designed to measure the constructs under investigation. The survey used forced 4-point Likert scale questions to elicit decisive responses from participants, thus reducing potential biases. Additionally, the survey incorporated a checking question to verify that respondents thoroughly read all questions before providing their answers, further ensuring data quality. Upon completing the data collection process, the team meticulously reviewed and processed the data to ascertain their quality and accuracy before advancing to subsequent data analysis.

Data Analysis
The data analysis for this study consisted of 2 primary stages: descriptive statistics and partial least squares structural equation modeling (PLS-SEM). Descriptive statistics were calculated for all survey questions to provide an overview of the responses’ central tendency, dispersion, and distribution. These statistics offered initial insights into the participants’ attitudes and perceptions regarding the constructs under investigation.

Following the descriptive analysis, the research team used PLS-SEM to examine the relationships between the latent constructs. PLS-SEM is a powerful multivariate analysis technique that allows researchers to estimate complex cause-effect relationships between latent constructs and their indicators [32]. This method was chosen for its ability to handle small- to medium-size samples and suitability for exploratory research [33]. The PLS-SEM analysis in our study was conducted in 2 stages: the assessment of the measurement model and the evaluation of the structural model. We assessed the measurement model for reliability and validity by focusing on 4 aspects: indicator reliability, internal consistency reliability, convergent validity, and discriminant validity. Indicator reliability was examined by analyzing the factor loadings of each indicator, with loadings greater than 0.5 considered satisfactory. We evaluated internal consistency reliability using composite reliability (rhoC), and values above 0.7 were deemed acceptable [34]. Convergent validity was assessed by examining the average variance extracted (AVE), and values above 0.5 indicated an adequate convergent validity [34,35]. In addition to these assessments, we also evaluated the reliability of the constructs in our research model using the average interitem correlation (rhoA). Both rhoC and rhoA are measures of internal consistency that help determine how closely related the survey questions are within each construct. A value of 0.7 or higher for both rhoC and rhoA is generally considered to indicate satisfactory reliability.

After confirming the measurement model’s adequacy, we evaluated the structural model to test our hypotheses. This analysis included assessing the path coefficients, significance levels, and determination coefficients ($R^2$) for each endogenous latent construct.

Results

Participant Details
A total of 607 individuals participated in the study, providing comprehensive responses to the questionnaire. Table 2 shows the descriptive statistics of the study variables. Most of the respondents used ChatGPT for information search and entertainment purposes. Others used the technology to solve problems and conduct health-related searches. Most respondents were willing to use ChatGPT for self-diagnosis (n=476, 78.4%). Most of the respondents were also familiar (to a certain degree) with the technology of ChatGPT and perceived the technology to be persuasive. Most respondents had a bachelor’s degree, a high school diploma, or a master’s degree.

Figure 2 illustrates the proportion of ChatGPT use frequency, respondents’ purpose of using ChatGPT, their familiarity with the technology, their perception of ChatGPT’s persuasiveness, and their education level.

Our study investigated the relationships between performance expectancy, risk-reward appraisal, decision-making, and the intent to use ChatGPT for self-diagnosis. The $R^2$ values indicated that our model can explain 52.4% of the variance in decision-making and 38.1% in the intent to use ChatGPT for self-diagnosis. When adjusting for the number of predictors in the model, the $R^2$ values were 52.2% for decision-making and 37.9% for the intent to use ChatGPT for self-diagnosis, demonstrating satisfactory explanatory power.

Regarding reliability, the performance expectancy construct had a Cronbach $\alpha$ coefficient of .783, a rhoC of .860, and an AVE of 0.606. The decision-making construct exhibited a Cronbach $\alpha$ coefficient of .668, a rhoC of .858, and an AVE of 0.751. The rhoA values were similar to Cronbach $\alpha$ values for each construct, further supporting the reliability of the constructs. Moreover, the AVE values for all constructs exceeded the recommended threshold of 0.5, indicating adequate convergent validity.
Table 2. Descriptive statistics of study variables.

<table>
<thead>
<tr>
<th>Question</th>
<th>Mean (SD)</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>PE1(^a)</td>
<td>3.24 (0.77)</td>
<td>0.694</td>
</tr>
<tr>
<td>PE2</td>
<td>3.22 (0.78)</td>
<td>0.285</td>
</tr>
<tr>
<td>PE3</td>
<td>3.20 (0.74)</td>
<td>0.290</td>
</tr>
<tr>
<td>PE4</td>
<td>3.24 (0.76)</td>
<td>0.403</td>
</tr>
<tr>
<td>DM1(^b)</td>
<td>3.25 (0.78)</td>
<td>0.633</td>
</tr>
<tr>
<td>DM2</td>
<td>3.13 (0.81)</td>
<td>0.028</td>
</tr>
<tr>
<td>RRA(^c)</td>
<td>3.20 (0.80)</td>
<td>0.371</td>
</tr>
<tr>
<td>IU(^d)</td>
<td>3.09 (0.85)</td>
<td>−0.180</td>
</tr>
</tbody>
</table>

\(^a\)PE: performance expectancy.  
\(^b\)DM: decision-making.  
\(^c\)RRA: risk-reward appraisal.  
\(^d\)IU: intent to use.

Figure 2. Illustration of the proportion of ChatGPT use frequency, respondents’ purpose of using ChatGPT, their familiarity with the technology, their perception of ChatGPT’s persuasiveness, and their education level. ChatGPT: Chat Generative Pretrained Transformer.

Our findings provide empirical support for all 3 hypotheses. We discovered that higher performance expectancy and positive risk-reward appraisals of ChatGPT are positively associated with improved decision-making outcomes among users. Additionally, enhanced decision-making processes involving ChatGPT positively impact users’ intention to use the technology for self-diagnosis. The results, including standardized β coefficients, SDs, t values, and 95% CIs, are presented in Table 3.
PLS-SEM analysis results elucidated the significant associations among the study variables. The direct effects indicated a significant positive association between performance expectancy and decision-making ($\beta=.547$, $t=14.715$) and between risk-reward appraisal and decision-making ($\beta=.245$, $t=5.850$). Moreover, the analysis identified a noteworthy positive relationship between decision-making and the intent to use ($\beta=.565$, $t=16.928$). Concerning indirect effects, the findings revealed that performance expectancy significantly influences the intent to use, mediated by decision-making ($\beta=.309$, $t=10.911$). Likewise, risk-reward appraisal demonstrated a meaningful positive impact on the intent to use via decision-making ($\beta=.138$, $t=5.191$).

**Table 3.** Standardized direct, indirect, and total effects.

<table>
<thead>
<tr>
<th>Effects</th>
<th>$\beta$ (SD)</th>
<th>$t$ value</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct effects of study variables</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance expectancy à decision-making</td>
<td>.547 (.037)</td>
<td>14.715</td>
<td>0.474-0.620</td>
</tr>
<tr>
<td>Risk-reward appraisal à decision-making</td>
<td>.245 (.042)</td>
<td>5.850</td>
<td>0.161-0.325</td>
</tr>
<tr>
<td>Decision-making à intent to use</td>
<td>.565 (.033)</td>
<td>16.928</td>
<td>0.498-0.628</td>
</tr>
<tr>
<td>Indirect effects of study variables</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance expectancy à decision-making à intent to use</td>
<td>.309 (.028)</td>
<td>10.911</td>
<td>0.255-0.366</td>
</tr>
<tr>
<td>Risk-reward appraisal à decision-making à intent to use</td>
<td>.138 (.026)</td>
<td>5.191</td>
<td>0.087-0.191</td>
</tr>
<tr>
<td>Total effects of study variables</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance expectancy à decision-making</td>
<td>.547 (.037)</td>
<td>N/A$^a$</td>
<td>0.474-0.620</td>
</tr>
<tr>
<td>Performance expectancy à intent to use</td>
<td>.309 (.028)</td>
<td>N/A</td>
<td>0.255-0.366</td>
</tr>
<tr>
<td>Risk-reward appraisal à decision-making</td>
<td>.245 (.042)</td>
<td>N/A</td>
<td>0.161-0.325</td>
</tr>
<tr>
<td>Decision-making à intent to use</td>
<td>.138 (.027)</td>
<td>N/A</td>
<td>0.087-0.191</td>
</tr>
</tbody>
</table>

$^a$N/A: not applicable.

**Discussion**

**Principal Findings**

We acknowledge that conversational AI systems, such as ChatGPT, can be crucial in health care by providing numerous possibilities to elevate patient care, optimize medical workflows, and augment the overall health care experience. In this study, we highlighted specific obstacles that must be tackled for secure and efficient implementation of ChatGPT. Although our study is the first to investigate the effects of perceived ChatGPT effectiveness and risk-reward appraisal on decision-making, its validity is supported by numerous studies examining the relationships among performance expectancy, technology acceptance, usage, risk-reward appraisal, and decision-making in other domains.

**Contributions of This Study**

The findings of our study contribute to the growing body of the literature on AI chatbots in health care and their potential applications, particularly in the context of self-diagnosis. In recent years, research has increasingly focused on developing and evaluating AI chatbots for various health care purposes [35,36]. However, our study is unique in that it specifically examines the factors influencing users’ intentions to use ChatGPT, an AI chatbot not designed for health care purposes, for self-diagnosis. This novel focus allows for a deeper understanding of users’ perceptions and behaviors in the context of AI chatbots and self-diagnosis, which can be crucial for ensuring the safe and responsible integration of such technologies into health care.

Our research builds on earlier studies investigating the factors affecting the adoption of AI chatbots in health care [37-40]. Although these studies have provided valuable insights into the factors driving the adoption of AI chatbots, our study extends this knowledge by examining performance expectancy, risk-reward appraisal, and decision-making processes as key determinants of users’ intentions to use ChatGPT for self-diagnosis. This nuanced analysis can help inform the design and implementation of AI chatbots in health care and also help develop policies and interventions to mitigate the potential risks of using such technologies for self-diagnosis.

By focusing on ChatGPT, our study contributes to the broader conversation on AI chatbots’ ethical and societal implications in health care. The increasing popularity of AI chatbots, such as ChatGPT, for self-diagnosis raises important questions about the responsibilities of AI developers, health care providers, and policy makers in ensuring such technologies’ safe and responsible use. Our findings highlight the need for a
collaborative, interdisciplinary approach to addressing these challenges, involving stakeholders from various sectors, including AI development, health care, policy, and ethics.

Implications

The implications of our findings from a policy and pragmatic standpoint suggest a need for proactive preparation and policy alteration concerning the use of ChatGPT for self-diagnosis in health care.

Human behavior has consistently demonstrated a tendency to repurpose technology for purposes beyond its original design, even when aware of the potential risks or drawbacks. In the context of our study, people are inclined to use ChatGPT, a technology not specifically designed for health care applications, for self-diagnosis, as they perceive it to be useful and easy to use. Similarly, as evidenced by our prior study on internet use and mental health, people often turn to online sources for self-diagnosis and health information, despite the potential negative impact on mental health [41]. The reliance on these sources can be attributed to “curiosity gap” theory [42], which suggests that individuals are motivated to seek information to reduce uncertainty, even when the information may not be entirely accurate or reliable. This drive for information, combined with the convenience and accessibility of technology, may result in people using tools like ChatGPT or the internet for self-diagnosis, despite their inherent limitations.

In both cases, people’s behavior can be understood by observing the balance between perceived benefits, ease of use, and potential risks. The desire to reduce uncertainty and the convenience of technology may outweigh the awareness of potential drawbacks or misuse. This highlights the need to develop and regulate technologies like ChatGPT and online health information sources to meet health care applications’ unique requirements and ethical considerations, ensuring that they are user-friendly and trustworthy and minimize negative impacts on users’ health and well-being.

First, policy makers and health care stakeholders should collaborate to establish guidelines and ethical standards for using ChatGPT in health care settings [43]. These guidelines should consider the potential risks, benefits, and limitations of using AI-powered chatbots in health care, such as patient privacy, the health care applications’ unique requirements, and the ethical considerations researchers should focus on, to enhance the performance, safety, and accuracy of ChatGPT for health care applications.

Second, by tailoring the chatbot to address medical inquiries and concerns better, users can receive more reliable and valuable information to inform their decision-making processes. In addition, incorporating evidence-based medicine, reliable sources, and expert opinions into the chatbot’s knowledge base can further improve its credibility and usefulness in the health care [44]. To extend this implication, an integrated diagnostics mechanism could be developed to enhance ChatGPT’s ability to assist users with self-diagnosis. This mechanism would involve combining various diagnostic tools and techniques, such as symptom checkers, medical history analysis, and even integration with wearable health-monitoring devices, to gather real-time data. ChatGPT could then analyze the information provided by these sources to generate more accurate and personalized assessments of the user’s health condition.

Third, educating and informing users about the appropriate use of ChatGPT for self-diagnosis and its limitations are essential. Public health campaigns and educational materials should emphasize the importance of consulting health care professionals for accurate diagnosis and treatment, while highlighting the potential benefits of using chatbots as an adjunct tool for health information and decision-making support. A feedback mechanism could be proposed to ensure shared understanding and improve user awareness. This mechanism would involve users providing feedback on their experience with ChatGPT, including the accuracy and relevance of the information received and any concerns or misconceptions they may have encountered.

Health care professionals could also be involved in this process, sharing their perspectives on the chatbot’s performance and suggesting improvements to enhance its reliability and user-friendliness. The feedback collected would then be used to refine ChatGPT’s algorithms, knowledge base, and user interface, ensuring it remains current with the latest medical knowledge and best practices. This iterative process would foster continuous improvement of the chatbot’s performance and promote greater awareness and understanding among users about the appropriate use of ChatGPT and its limitations in the context of self-diagnosis. Additionally, educational resources, such as tutorials and guidelines, could supplement the feedback mechanism to guide users in interacting with ChatGPT effectively and responsibly. By implementing a feedback mechanism and providing educational support, users can better perceive ChatGPT’s capabilities and limitations, ultimately promoting responsible and effective use of AI chatbots in health care settings.

Lastly, continuous monitoring and evaluation of ChatGPT’s use in health care should be conducted to assess its impact on health care outcomes and decision-making. This will enable policy makers and health care providers to make informed decisions about the potential benefits, risks, and practical applications of ChatGPT in health care settings.

Limitations

Our study has limitations that warrant consideration. First, we did not control for potential confounding factors, such as age, medical condition, health literacy, previous experience with comparable technologies, or demographic characteristics, which might significantly influence users’ intentions to use ChatGPT for self-diagnosis. The results among younger and healthier populations could differ substantially from those among older populations with existing medical conditions. Younger individuals may be more inclined to use AI chatbots due to their familiarity with technology. In comparison, older individuals or those with medical conditions may seek additional reassurance or support for managing their health concerns.

Second, the cross-sectional survey design constrained our capacity to examine the evolving nature of users’ interactions with AI chatbots. Moreover, relying on self-reported measurements may introduce various biases, including social desirability, recall, or imprecise reporting. Self-report measures
obtained through surveys inherently capture users’ perceptions rather than objective reality. Although the participants’ subjective experiences can provide valuable insights, there may be discrepancies between these perceptions and the actual situation. Furthermore, the cross-sectional design of the study limited our ability to draw causal inferences over time. Future research could use a triangulation approach to mitigate these limitations, incorporating objective measures and longitudinal data collection to provide a more comprehensive understanding of the phenomenon under investigation. Lastly, focusing on ChatGPT, which is not specifically intended for health care applications, may limit the applicability of the findings to other AI chatbots explicitly designed for health care purposes.

To address these limitations, future research should consider using longitudinal data, stratifying the sample by age group and medical condition, and accounting for potential confounding factors, such as participants’ familiarity with AI technology, prior experiences with chatbots, and demographic information. Various methodologies could provide additional insights, including monitoring chatbot usage and conducting qualitative interviews to assess trust and user behavior. Enhancing the data collection frequency and guaranteeing participant anonymity may also help reduce biases. By addressing these constraints, future research can contribute to a more comprehensive understanding of AI chatbot adoption in health care settings and enable more targeted interventions to optimize patient care and outcomes across diverse populations and health statuses.

Conclusion
In conclusion, our research investigated the factors influencing users’ intentions to use ChatGPT for self-diagnosis, a purpose for which the technology is not specifically designed. The study aimed to explore the implications of these factors for the safe and effective integration of AI chatbots in health care settings. By examining performance expectancy, risk-reward appraisal, and decision-making processes, our findings contribute to the growing body of the literature on AI chatbots in health care and provide insights into AI chatbot adoption in health care contexts.

The clinical message of this study is to emphasize the importance of ongoing collaboration among AI developers, health care providers, and policy makers in ensuring the safe and responsible use of AI chatbots in health care. Addressing users’ expectations, risk-reward appraisal, and decision-making processes can help develop AI chatbots tailored to human needs and preferences, providing consumers with reliable and verified sources for health-related information. This approach can not only enhance health care accessibility but also improve health literacy and awareness among the public.

As the field of AI chatbots in health care continues to evolve, future research should further investigate the long-term effects of using AI chatbots for self-diagnosis and explore the potential integration of AI chatbots with other digital health interventions to optimize patient care and outcomes. In doing so, we can better understand the implications of AI chatbot usage in health care settings and ensure that these technologies are designed and implemented to safeguard users’ well-being and support positive health outcomes.

Authors’ Contributions
AC, the lead researcher, was responsible for the study’s conceptualization, the survey’s development, figure illustration, data collection, and manuscript writing. YS, the student author, contributed to the study by conducting data analysis and providing input during manuscript writing. Both authors equally collaborated throughout the research process, ensuring the rigor and accuracy of the study, and approved the final version of the manuscript for submission.

Conflicts of Interest
None declared.

References


23. Shahsavar & ChoudhuryJMIR HUMAN FACTORS


Abbreviations

AI: artificial intelligence
AVE: average variance extracted
ChatGPT: Chat Generative Pretrained Transformer
e-commerce: electronic commerce
PLS-SEM: partial least squares structural equation modeling
PMT: protection motivation theory
TAM: technology acceptance model
UTAUT: unified theory of acceptance and use of technology

©Yeganeh Shahsavari, Avishek Choudhury. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 17.05.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.